PhD-FHSE-2024-048
The Faculty of Humanities, Education and Social Science

DISSERTATION

Defence held on 19 December 2024 in Esch-sur-Alzette
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN GÉOGRAPHIE

by

LUCAS VIEIRA MAGALHÃES
Born on 6 May 1994 in Belo Horizonte (Brazil)

PATTERNS AND DYNAMICS OF RETAIL FIRM LOCATION

Dissertation defence committee

Prof. Dr Geoffrey CARUSO, Dissertation supervisor
Professor, University of Luxembourg

Prof. Dr Pierre PICARD
Professor, University of Luxembourg

Dr Philippe GERBER
Luxembourg Institute of Socio-Economic Research

Dr Justin DELLOYE
Université Catholique de Louvain

Dr Marion LE TEXIER
Université Paul Valéry Montpellier 3

Acknowledgements

I am deeply grateful to everyone who supported me throughout this journey, especially during the challenging periods of social distancing. To my family in Brazil for always providing love and support, online when face-to-face was not an option. To my friends in Brazil, who stayed close despite the distance. To the Brazilian friends with whom I now share this journey in Europe. To the new friendships I have formed in Luxembourg, who have made this place a true home away from home, especially through the loneliness of European winters. To my friends in Uni and LISER, especially the mezzanine crew, for brightening the day-to-day with good humour, lunch talks, after-work gatherings, ping-pong and, of course, inspiring conversations and academic support.

I am also especially grateful to my supervisory team for their guidance, patience, and encouragement throughout this challenging process. And to the ACROSS team, first for entrusting me with their confidence, but above all for the provided opportunities and the hard work towards fostering an environment of academic excellence.

To all those who have been part of this journey, in ways both big and small, I extend my heartfelt thanks. This work stands as much a reflection of your support as of my own efforts.

This dissertation was developed at the Luxembourg Institute of Socio-Economic Research (LISER) in partnership with the University of Luxembourg under the Doctoral Training Unit ACROSS - Analysis of CROSS-border human mobility, funded by the Luxembourg National Research Fund (FNR) (PRIDE19/14302992).

Contents

List of Figures

List of Tables

Abstract

Shopping has long been central to urban life, with cities serving as places where demand meets supply for both everyday and specialised goods. However, recent trends in shopping habits (e.g., online shopping), working patterns (e.g., teleworking), and urban development policies (e.g., the 15-minute city) have begun to challenge traditional push-and-pull forces and alter retail firm locations, agglomeration patterns, and consumer behaviour. This dissertation aims to understand how and where shops cluster, and how these patterns evolve as cities grow and develop, providing insights into the complexifying interplay between consumption and urban form. The dissertation is divided into two parts: the first examines observed retail patterns, and the second analyses the dynamics behind these patterns. In the first part, I i) extract retail location data from OpenStreetMap (OSM) for 782 European cities, showing that shops scale super-linearly with city size, while categories scale sub-linearly; ii) analyse intra-urban patterns, calculating walking-distance accessibility under the 15-minute city framework, highlighting that middle-sized Southern European cities are particularly accessible; and iii) assess intra-urban clustering of economic activities, applying the Economic Complexity framework. In the second part, I iv) introduce a theoretical urban model to assess the viability of central retail under online shopping and teleworking preferences; and v) implement this model in an agent-based simulation to explore different location patterns as firms and households interact in space. The results contribute to theoretical and methodological discussions in the field, highlighting how 1) larger cities may offer disproportionately more shops, but middle-sized European cities offer higher accessibility to retail; 2) specialised goods strongly depend on centrality to be viable; 3) consumption generates agglomeration, and strong agglomerations remain key drivers of retail viability, meaning normative models like the 15-minute city must

account for these dynamics; 4) both teleworking and online-shopping behaviours act towards less viability of central retail; 5) century-long theoretical discussions (e.g., minimum vs. maximum differentiation) may be reconciled in an agent-based simulation. A proper understanding of the future of urban retail and its relationship to changing consumption trends requires a detailed analysis of existing patterns and a simplified abstraction of complex dynamics, which this thesis attempts to bring to light.

The chapter-specific titles and abstracts are presented below:

Chapter 1 - The scaling of retail in European cities

This paper presents an analysis of the spatial distribution and diversity of retail establishments across 782 European cities, using data from OpenStreetMap (OSM). By examining both the total number of shops and their categorical composition, we investigate scaling patterns to understand how the retail ecosystem evolves with city size. Our findings reveal that larger cities exhibit a disproportionately higher quantity of retail establishments (superlinear scaling with an exponent of 1.14), while the total number of categories and their diversity increase positively but sublinearly with population. We observe a decreasing scaling coefficient for individual retail categories as their ubiquity decreases, indicating that only the most ubiquitous (common) goods exhibit superlinear scaling. This pattern suggests that the overall superlinearity in retail scaling is primarily an inter-sectoral phenomenon resulting from the aggregation of multiple categories. Additionally, we explore the nestedness of retail categories and find a high degree of nested hierarchical structure (NODF = 0.78), confirming the hypothesis derived from Central Place Theory (CPT) that smaller cities offer subsets of the more diverse retail environments of larger urban areas. Our results demonstrate that the scaling of retail in European cities is driven by a combined effect of diversification and specialisation, with common goods playing a dominant role and rarer goods collectively enhancing the superlinear scaling through their aggregated presence. These insights provide a valuable baseline for further comparative studies on urban retail landscapes.

Chapter 2 - The *retailscape* of European cities: assessing the 15-minute accessibility of retail location with open data

Retail constitutes a key function of urban life. Shops not only connect demand to supply of both everyday and specialised goods, but they also serve as an important provider of amenities for citizens, especially with the rise of online-shopping. This study analyses retail accessibility across 782 European Functional Urban Areas (FUAs) using Open Street Map (OSM) data for retail location, and the Global Human Settlement Layer (GHSL) for population data. We analyse spatial patterns of retail distribution across cities and compare them to local population density to highlight local and regional differences in accessibility to retail. Firstly, we calculate travel times between intra-urban hexagonal grid cells and shops, accounting for space frictions using the concept of accessibility. Then, we calculate, for each FUA, a series of time-bounded accessibility metrics and a population-weighted city-level accessibility indicator, allowing for cross-city comparisons and country-level aggregations. Lastly, we propose a *retailscape* tightness measurement by comparing two time-bounded accessibilities, under the lens of the 15-minute city concept. Findings reveal that middle-sized Mediterranean cities, especially in Spain and Greece, exhibit the highest accessibilities to commerce within shorter travel times, while larger metropolitan areas like Paris, London, and Berlin dominate longer thresholds. Finally, we highlight that regions of Spain, Southern Italy, Albania and Greece exhibit a highly tight *retailscape*, approximating them to the model advocated by the 15-minute city. This study focused on observed patterns of retail distribution, encouraging further research to investigate specific mechanisms that make these patterns emerge.

Chapter 3 - Bringing economic complexity to the intra-urban scale: the role of services in the urban economy of Belo Horizonte, Brazil

This study explores the formation of economic complexity within a city from the Global South, during 2011-2019. It proposes an expanded interpretation of the Economic Complexity Index (ECI) to be applied at the intra-urban context of Belo Horizonte, Brazil, focusing on three different spatial levels of analysis (i.e., local, neighbourhood, and community levels). By applying the index to these three levels, instead of regional or national administrative boundaries commonly used in literature, this study contributes to approximating the observation of economic complexity to the actual geographical scales at which economic interactions take place, allowing for intra-urban comparisons. The proposed ECI includes the service economy, amenities, and retail, in addition to commonly observed manufacturing industry. Methodologically, this case study introduces the Urban Economy Space network diagram to the expanded ECI as an effort to holistically consider all economic sectors happening in a city. The main findings are twofold. First, the city services classified as more complex by the ECI are aligned with the theory of post-industrial economic activities: financial, telecommunications, scientific and technical services, etc. Second, government-led institutions such as healthcare facilities, higher education institutions, etc., appear on the top tier of economic complexity, indicating that local and national governments can contribute to complexifying local economies.

Published in Applied Geography (2023) with co-authors Monika Kuffer, Nina Schwarz, Monica Haddad.

Chapter 4 - Digital or local? Impacts of online shopping and teleworking on urban expansion: a theoretical approach

The ongoing shift towards online shopping, a trend further accelerated by the Covid-19 pandemic, has impacted urban spaces. Coupled with the changes in working habits, centrally located retail and office spaces might be severely affected. Despite these disruptive trends, the potential drop in rental and property prices, resulting from decreased urban demand, could revive attraction towards city centres, generating conflicting forces that need to be further studied. In order to address these issues, this paper proposes a simplified linear city model. Departing from the classic monocentric city model, it aims to emulate the interactions between households and firms, contributing to the discussions around the future of consumption, retail spaces, and urban form. By examining the bid-rent curves generated by different consumption behaviours, the paper explores how the interplay between various factors such as online shopping behaviour, import costs, and frequency of shopping and commuting trips can shape the built-up landscape of a fictitious city. Analysis of the equilibrium suggests that different consumption behaviours might directly affect the viability of central retail and even disrupt urban fabric, indicating the importance of the resilience of local governments in the face of changes in working habits and consumer behaviours.

Chapter 5 - Simulating retail location in the digital era: a spatially-explicit agent-based monocentric city model

Retail location competition is further complexified in the digital age, with the establishment of competitive online shopping alternatives. This study proposes a spatially-explicit agent-based model (ABM) to explore retail location dynamics

in a two-dimensional monocentric city framework, integrating classical economic theories and recent challenges, such as digital commerce. We show that firms' location choices and economic viability are significantly influenced by the interplay between importing fees, income levels, and the opportunity cost of land. The model also bridges the principles of minimum differentiation, as proposed by Hotelling (1929), and maximum differentiation, by d'Aspremont et al. (1979), testing emergent patterns under varying transportation cost structures - linear and exponential. Findings reveal that both minimum and maximum differentiation can coexist, even in a pure exponential transportation cost implementation - where pure maximum differentiation would be expected otherwise. The study contributes methodologically to the field of location competition and urban economics by incorporating agent-based simulations to capture the complexities of retail location decision in a dynamic urban environment, where a large number of agents (households and firms) interact.

Introduction

> When distance and convenience sets in; the small, the various and the personal wither away.
>
> *Jane Jacobs*

Cities have long been places of exchange. Since ancient times, it has been where citizens have met to exchange goods, from market squares of medieval Europe to bazaars in the Middle East. Producers and consumers have traditionally traded in central areas, often mediated by specialised sellers. It is, effectively, where demand meets supply, for both everyday and specialised goods. With time, consumption spaces have reflected both the cultural and economic tendencies of their period. Not only defined by the traded products themselves but the very environments in which this trading has taken place. From square markets to shopping streets; from galleries to, eventually, shopping malls, much of a society's characteristics are reflected in how and where it consumes. Consumption habits, behaviours and preferences are, thus, deeply intertwined with the physical characteristics of the places of exchange in a way that understanding the latter can offer a lens to the functioning of a society at large.

In Europe, the 19th century brought along profound changes in production and societal structures. With the advent of industrialisation, the recently consolidated upper classes, along with an emerging *petite bourgeoisie*, found a way to demonstrate their wealth to their peers via consumption. In a new logic where lordships and land titles were losing importance in favour of wealth and possessions, ever-differentiating consumer goods gained momentum, with luxury clothing and jewellery being exhibited in balls and *promenades*. In the *Belle Époque*, the opening of boulevards and shopping galleries aligned the leisure activities of this new society to this consuming culture and the actual development of new consumption spaces. Shopping galleries such as Milan's Galleria Vittorio Emanuele II combined the fashionable *art-nouveau* architectural style with the novel steel-and-glass engineering works. Spaces dedicated to retail were becoming icons of the new technologies in construction being exhibited at World Fairs all over the globe. In a similar fashion, multi-product department stores following the same architectural trends were being opened all over Europe: the Galeries Lafayette in Paris, the Harrods in London, and the Bijenkorf in Amsterdam are a few examples. Consumption spaces were never mere places of exchange but where society would come together to also exchange experiences and influences.

By providing amenities and opportunities in this new thriving environment, cities ended up attracting an intense influx of dwellers originating from the countryside. Rising incomes, labour rights, and improving working conditions allowed an incipient middle class to also participate in this game of appearances. The development of the assembly line and scientific management allowed consumer goods to be popularised, kick-starting what would be later called consumerism. Initially restricted to textiles, it then expanded to manufactured goods, powered by interchangeable parts and steam machines. The division of labour and the ever-specialising processes came to dominate most aspects of production and economic life at large, until these days. Perhaps the most prominent example of this was the private automobile. Ford's Model T is generally regarded as the first mass-affordable private automobile, released in 1908. Fordism is an illustration of how the new ways of mass production allowed for the emergence of a salaried middle class, who could in turn afford the products produced *en masse*, generating a self-reinforcing cycle.

However, the intense centripetal forces attracting population to the cities also caused problems. Overcrowding and congestion soon came to place, especially in sanitary and housing terms. The attempt to solve these diseconomies of agglomeration originated a set of urban utopian theories, aiming at *exiting* the city, claiming that it would be possible to develop a fundamentally higher quality of life in self-contained communities out of main centres. Prominent examples of these utopias are Fourier's *phalanstères* and Ebenezer Howard's Garden City. These normative models promised to combine sanitary benefits from the rural, bucolic countryside with the level of services cities could offer. They came to influence a wide range of subsequent urban theories and planning practise in the coming years, leading even to cities built from scratch following these ideals.

Together with the development of communication technologies (e.g., the radio, the telegraph), the advances in transportation modes (e.g., the automobile, railway systems) gave rise to another set of urban theories regarding the “death of distance”: i.e., people, goods, and, eventually, information would be so easily transported that distance would not matter anymore. A conceptual urban model incorporating this thought is Frank Lloyd Wright’s Broadacre City (1932), in which households would live in self-sustainable estates (of one acre of land) while travelling freely between acres to interact with one another. This extreme decentralisation and suburbanisation, of course, would only be possible if the use of the private automobile was pervasive.

Over the course of the 20th century, the ownership and usage of the private car rose to dominate everyday life and fundamentally reshaped lifestyles and the cityscape in most of the Western world. The expansion of highway systems allowed residents to suburbanise, avoiding the feeling of overcrowdedness in traditional city centres. By following demand, consumption spaces have also suburbanised, especially in the post-war: the shopping mall emerges as a suburban concentration of shops, accessed by cars in previously remote areas. The modernist push towards land-use segmentation has further intensified this trend, with zoning ordinances determining which uses would be allowed where. Modernist urbanism was also influenced by the self-contained communities of decades before, proposing concepts such as the neighbourhood-unit, in which these rational cities needed to provide essential services to inhabitants of their vicinity. The overall idea that the alleged chaos of traditional centres could be overcome by an overarching rationality of city planning was dominant in the modernist thought, and intensely present in planning practise, especially in the post-war.

The overrationality of the modernist city did not come unquestioned. Perhaps the biggest exponent of this criticism was Jane Jacobs, who claimed in the 1960s that the infrastructure dedicated to the automobile, central in the urban planning of the time, was fundamentally killing urban life. She claimed that the very point of urbanity was the diversity of uses, people, and economic activities that only cities could offer. Still, rational, top-down interventions on the urban fabric were hammering these benefits, dilacerating existing communities. Proximity to services and amenities, as a fundamental urban quality, was also put forward by more recent urban theories, such as the "New Urbanism" movement in the 1990s and the "15-minute city", in the 2020s. The latter reappropriated the previously explored concept of the self-contained communities, in which the average urban dweller would find an amalgam of services, amenities and - as a matter of fact - shops, within 15 minutes from their household.

Nonetheless, a fundamental contradiction remains. In an ever-globalising world, where the distance burden — also called the friction of space — is, indeed, diminishing, the idea of a self-contained community can become nothing but a panacea. In other words, reducing spatial frictions would make self-contained communities heavily connected, contradicting their very concept. This is especially the case when the manufacturing industries, initially spinning the wheel of Fordism in the global North, are being outsourced to the global South in a search for reduced labour costs and regulations. The current consumerism prevailing in Western society is a deeply interconnected one, heavily dependent on production happening elsewhere, and consumer goods nowadays essentially travel long distances before they hit the shelves of stores. In some cases, they might not even hit the shelves at all, as online-shopping has also become prevalent.

The Covid-19 outbreak reinforced these hyperconnected consumption trends. During the pandemic, consumers shifted their preferred shopping places from local shops and malls towards internet-based orders, especially via the smartphone, with an increase in multi-brand applications and standard retailers investing in mobile-optimised websites and apps. These were not recent trends, but diminishing transportation costs for decades and increased logistics in short distances (the so-called last-mile problem) made it increasingly viable. Recently, online shopping has reached even grocery shopping, which was previously more resistent to this phenomenon. During the pandemic itself, consumers avoiding overcrowded shops helped catapult this habit even further.

Despite being triggered by a specific moment, some of these changes are expected to last in the long run. The popularity of smartphones is nowhere near its peak, and together with it comes the intensification of mobile shopping. Coupled with an increasing capacity of social network algorithms to personalise the content consumers are bombarded with, ever-more personalised online shopping experiences also contribute strongly to this trend. In terms of grocery shopping, with the logistic infrastructure, both physical and digital, installed and growing, it also seems like a trend that is here to stay.

These phenomena may have significant impacts on cities and urban floor-space. Traditional brick-and-mortar retailers are already facing declining demand from by-passers, leading to a crisis of the so-called “high-streets”. Former centrally located retail floor-space is now under pressure towards being repurposed to other uses, including even temporary accommodation and hospitality. The concomitant rise in restaurant delivery services gave rise to “dark kitchens”, while online grocery shopping to “dark stores”, intensively using floor-space on ground-floors in centrally located areas but not open to the public, nor fluid to pedestrians. On the other hand, suburban floor-space, especially for logistic purposes, may be on the rise. Where the land is cheaper, it is easier to install up-to-scale warehouses and distribution centres that support the whole online-shopping consumption patterns.

Additionally, the shifting tendency towards teleworking also contributes to accelerating these changes. Alternatively, the pressure is more strongly felt on vertical, out-of-the-ground spaces: parallel reducing demand for centrally located offices and urban dwellings, resulting in a shift away from urban areas. Again, these are not recent discussions, as the idea of being able to work from anywhere has always been around, especially when paradigm shifts in transportation have come into place. However, the fundamental contradiction mentioned before - between push & pull forces - may lead our cities towards not-so-obvious paths. A “general equilibrium effect” can be observed when talking about an urban flight: reduced demand for central areas may trigger a downward adjustment in rental and property prices, making them attractive again to certain parcels of the population or to certain economic sectors, which can again reignite demand for central areas. This was already observed recently with the flight from the centres since the 1980s - especially due to rising criminality - and the recent reignition of central areas since the 2000s. More recently, the shifts in consumption have also given way to the questioning of their environmental sustainability, as long-haul delivery chains have a strong environmental impact. This led to a call towards buying locally, promoting local economic opportunities while valuing local products and processes when consuming. Of course, these ups and downs are worth observing and studying by themselves, as the incoming public may be very different from the outgoing one, a phenomenon that was conventionally called "gentrification" and is the object of several academic works.

How these forces interact, especially with recent shocks and trends, is uncertain. More specifically, the magnitude and timing of these adjustments, as well as its impacts on the urban form, depend on a number of factors, including local economic conditions and how governments will react to these challenges. Moreover, it is expected that some of these changes will have a more permanent facet, possibly leading to long-run spatial patterns for cities worldwide, likely never observed before. Considering holistically the pressures on the real-estate market, consumption behaviours from households, and selling strategies by retail firms may help planners, institutions, and businesses to adapt their courses of action to better direct these changes towards scenarios where greater societal welfare is the goal.

With all this in mind, this dissertation is presented to the academic community as an attempt to shed light on these dynamics and contribute to the discussions connecting the future of consumption, retail spaces, and cities in a broader sense. It does so by trying to build a bridge between literature in urban economics and quantitative geography, incorporating innovative methodologies such as agent-based simulations. This is done in attempts to extrapolate traditionally used analytical methods of urban economics, allowing us to account for non-linear dynamics of complex systems such as a city. On the other end, it enriches the urban modelling literature coming from quantitative geography, as it proposes microeconomic foundations for the interactions between agents.

This dissertation is composed of 5 chapters, each as a stand-alone research article. It is divided into two parts: one focusing on observed patterns, with 3 chapters; another focusing on exploring dynamics, with 2 chapters. In the first part, the first chapter presents the analysed dataset of location of retail firms for 782 European cities, exploring mainly scaling laws with population size. The second chapter dives into the intra-urban configuration of these locations, exploring walking-distance accessibility to these shops, aiming at assessing the ease of access to retail across these cities. The third chapter is a published article, deriving from the author's master's thesis, expanding the scope towards all economic activities in a city, and exploring measurements of economic complexity and diversity in a city of the Global South (Belo Horizonte, in Brazil), as a case-study. It is included in this dissertation for having also made use of accessibility indices, and exploring intra-urban clustering patterns of firms. In the second part, the fourth chapter proposes a microeconomic-based monocentric-city model with the added dynamics of online-shopping and teleworking, testing the viability of central retail under different conditions. Finally, the fifth chapter brings this model to a two-dimensional landscape, using agent-based modelling to derive spatial patterns of retail location and concentration, exploring how varying parameters related to consumption preferences and transportation behaviour affects the outcome patterns.

Chapter 1

The scaling of retail in European cities

This paper presents an analysis of the spatial distribution and diversity of retail establishments across 782 European cities, using data from OpenStreetMap (OSM). By examining both the total number of shops and their categorical composition, we investigate scaling patterns to understand how the retail ecosystem evolves with city size. Our findings reveal that larger cities exhibit a disproportionately higher quantity of retail establishments (superlinear scaling with an exponent of 1.14), while the total number of categories and their diversity increase positively but sublinearly with population. We observe a decreasing scaling coefficient for individual retail categories as their ubiquity decreases, indicating that only the most ubiquitous (common) goods exhibit superlinear scaling. This pattern suggests that the overall superlinearity in retail scaling is primarily an inter-sectoral phenomenon resulting from the aggregation of multiple categories. Additionally, we explore the nestedness of retail categories and find a high degree of nested hierarchical structure (NODF = 0.78), confirming the hypothesis derived from Central Place Theory (CPT) that smaller cities offer subsets of the more diverse retail environments of larger urban areas. Our results demonstrate that the scaling of retail in European cities is driven by a combined effect of diversification and specialisation, with common goods playing a dominant role and rarer goods collectively enhancing the superlinear

scaling through their aggregated presence. These insights provide a valuable baseline for further comparative studies on urban retail landscapes.

1.1 Introduction

The study of urban retail landscapes is crucial for understanding the economic and social dynamics of cities. Retail establishments not only serve as economic engines but also contribute to the vibrancy and livability of urban areas. With the proliferation of open data sources like OpenStreetMap (OSM), researchers now have access to detailed geospatial information on shop locations. Understanding how the quantity and composition of retail varies with city size is fundamental for planning for the future of cities. This paper uses OSM data to conduct a comprehensive analysis of retail distribution across 782 European cities, aiming to uncover scaling patterns and trends that can inform urban planning and policy-making.

Understanding the scaling laws that govern a system of cities is fundamental for making sense of how cities evolve with population growth. While certain phenomena tend to benefit from agglomeration, growing more than proportionately with population - e.g., employment, patents, as well as crime, pollution; others tend to scale sublinearly, by reaching either some sort of saturation or distributed costs - e.g., infrastructure, occupied areas (Ribeiro & Rybski, 2023). These are referred to as, respectively, socio-economic scaling (superlinear) and infrastructure scaling (sublinear). The retail ecosystem, despite being characterised in the literature as a complex system (Pennacchioli et al., 2014), has, to the best of our knowledge, not had its degree of scaling over a system of cities assessed. By benefitting from agglomeration economies, either in the form of knowledge spillovers, shared resources, or having access to specialised labour pools, retail systems are expected to follow the socio-economic scaling (superlinear). This brings us to our first and main research question: how does retail in a well-defined system of cities, such as the European one (Bettencourt & Lobo, 2016), scale with population?

Despite our main hypothesis—that retail does scale superlinearly with population—disentangling the forces behind it remains fundamental. Larger cities would tend to provide not only more goods but also goods that are not present in smaller ones. Either by being the places where innovation takes place (Jacobs, 1970; Pumain, 2011), or by being more centrally positioned in an interconnected network of demand (Verhetsel et al., 2022). The Central Place Theory (Christaller, 1966) provides a foundational framework for understanding this phenomenon. It states that more specialised, sophisticated, goods and services are expected only in more central places, while less specialised goods are more widespread. As a tool to fundamentally classify the level of centrality of a place, and of specialisation of a product or service, it has been widely used in retail location models - see Brown (1993). More recently, it has been reconciled with economic complexity theory by Kim et al. (2024).

The idea that a system is organised in a hierarchical fashion, where rare elements are necessarily observed in units of analysis with the more ubiquitous elements, brings us to ecological studies, or more specifically, studies of distributions of species in their habitats (Atmar & Patterson, 1993). A system is said to be *nested* when this hierarchisation happens, and the degree to which this is observed is defined as *nestedness* (Almeida-Neto et al., 2008; Jonhson et al., 2013). Similarly to ecosystems, retail categories as species and cities as habitats form a bipartite network, instrumental for detecting nested patterns and calculating the degree of nestedness. This brings us to our second research question: how nested are retail categories in European cities? Some degree of nestedness is expected by larger cities encompassing retail categories found in smaller cities, plus additional ones. This pattern would reflect larger cities' central roles in urban systems, the benefits of agglomeration economies, and the diffusion of innovations.

As larger cities are expected to not only have disproportionately more shops, but also host more specialised goods, it is fundamental to disentangle the composition of this scaling law and reconcile it with Central Place Theory. Studies like Arvidsson et al. (2023) highlight that within-city inequalities and the heavy-tailed distributions of human networking and productivity significantly contribute to different urban scaling effects, suggesting that different retail categories may scale differently based on their characteristics. This brings us to our third and last research question: what drives the scaling of retail in European cities? We propose to test the scaling of the different categories, as well as the scaling of diversity itself, to understand what forces prevail: either diversification (heterogeneity of goods) or specialisation (more of the same good).

The main objective of this research is, then, to investigate how the number of retail establishments and their categorical composition vary with city size across European cities. By examining the quantity and variety of retail outlets, we aim to identify geographical heterogeneities and regional differences, to understand how these factors vary across different urban environments. This analysis will provide insights into the economic composition and consumer behaviour in various regions, highlighting areas of retail concentration and diversity.

A secondary objective is to understand the hierarchical structure of shop categories across a system of cities of varying sizes. We test whether and to which degree this structure follows a nested pattern, in which the categories found in smaller cities are necessarily found in larger ones. Within the sphere of categorical composition, we also explore its diversity, by calculating both the number of categories present (richness) and their entropy per city.

Finally, we test how quantity, richness and entropy of shops scale with city size. Larger cities are expected to offer not only more shops, but a more diverse set of categories than smaller ones. However, whether these scale super- or sublinearly with city size is not trivial. Moreover, whether more ubiquitous or rarer goods drive this scaling is also unknown. Understanding these relationships is vital for urban economists and planners, as it can reveal how retail systems evolve with city growth and inform strategies for sustainable urban development.

In summary, this paper seeks to bridge a gap between scaling laws, ecosystem analysis, and Central Place Theory, by utilizing OSM data to explore retail landscapes across a diverse set of European cities. Through quantitative methods, we aim to provide a base for further analysis on how retail distribution and diversity are influenced by local geographical and demographic factors. We propose to, innovatively, 1) quantify the scaling of retail; 2) test the nestedness of retail categories across cities; 3) provide a specialisation ranking for retail categories; 4) disentangle scaling for different retail categories, according to their level of specialisation. The findings of this study may inform urban policy, economic development, and the future planning of retail infrastructure in European cities.

1.2 Definitions and methods

This section starts by defining the spatial units of this analysis, the data sources and the strategies used to extract and filter them. We then move to the metrics calculated for each spatial unit, namely the total quantity of shops, the richness of shop categories, the diversity of shop categories, and, finally, the nestedness temperature for testing the hierarchical composition of shop categories between cities. Finally, we describe how we calculate the scaling exponents for each of the metrics against city size.

1.2.1 Data definition

This section describes the datasets used for the analysis. We begin by identifying the main units of analysis, the Urban Atlas' Functional Urban Areas (FUA) of Europe. We then move to describe the extraction of population data from the Global Human Settlement Layer (GHSL). Finally, we describe the data extracted from OpenStreetMap (OSM), followed by the strategies to filter it.

We acquire the FUA's 2018 boundaries from the European Environmental Agency's Datahub. These include 788 cities and their commuting zones, whose labour markets are highly integrated with the main cores. They encompass all of the EU-27 countries, plus the UK, Switzerland, Iceland, Norway, the West Balkan countries and Türkiye. We use the boundaries of the FUAs to extract population data from GHSL's (Schiavina et al., 2023) 2020 100m-resolution population raster.

We extract OSM data on January 17th 2024 from Geofabrik Downloader for the whole of Europe. We choose to extract the data for the whole continent, and later clip it by the geographical units' boundaries for one main reason: the boundaries of Urban Atlas' FUAs do not necessarily coincide with the administrative units OSM has in record (i.e., sometimes FUAs include multiple municipalities which are not aggregated in OSM's "boundaries" definition). We converted all boundaries to GeoJson and used Osmium Command-Line Tool (Topf, 2013) to batch-clip OSM data for each FUA boundary. By doing that, we ended up filtering out 6 overseas territories, resulting in 782 geographic units for the analysis.

With OSM data separated for each FUA, we import them into the *python* environment using Pyrosm package (Tenkanen, 2020), which allows us to filter for the features that interest us. The keys and values used for filtering the dataset are shown in Table 1.1. In case polygons were returned (i.e., building footprints), they were converted to points using their centroids. In cases where multiple features were returned (e.g., the same shop accounted for in both "shop" and "building"), we take their unique IDs as identifiers and filter out repeated entries, making sure no duplicate entries are included.

Table 1.1: OSM's keys and values used for filtering the dataset.

Key	**Value**
'shop'	*True*
'building'	*'retail'*

Other attempts of filtering were also attempted, but excluded from this analysis. They were: 1) $'building' =' commercial'$, excluded from the analysis for including office buildings; 2) $'landuse' =' retail'$, excluded from the analysis for including aggregated shopping areas, such as malls, instead of the retail outlets themselves.

Overall, the total number of shops extracted is 1 540 995, spread over 4 216 unique business categories. To maintain inter-city consistency and avoid non-relevant categories, we also decided to keep only categories found in at least two cities. This reduced the sample to a more meaningful subset of 423 categories, keeping 1 534 165 total shops (or 99.6% of the initial total).

1.2.2 Metrics

As a first step we propose to conduct descriptive statistical analyses of the data collected and pre-processed so far. Firstly, we plot two maps to visualise the total number of shops per city, and the calculated population per city. This is regardless of the shop categories, and this total number of shops is also used for the scaling analysis described in the next subsection.

Secondly, we define two key concepts to depart from: diversity and ubiquity. A city's diversity refers to the total number of categories existing in it, while a category's ubiquity refers to the total number of cities in which it exists. Thus, diversity, defined as the absolute number, is akin to the concept of richness used in ecological studies, and defined here as K_c. It is a simple count of how many varieties are available to its citizens, regardless of the composition of this variety. On the other hand, ubiquity is category-specific, and is defined as the proportion of cities in the dataset containing that category. It is, then, defined as $u_k \in [0, 1]$. By referring back to Central Place Theory, we identify ubiquity as a fundamental concept to rank retail categories in terms of how specialised they are. In other words, the less ubiquitous categories (rare goods) are classified as the most specialised, by definition, while the most ubiquitous ones (common goods) as the least specialised.

The concepts of *ubiquity*—the number of cities in which a particular retail category is present—and *diversity* or richness—the number of retail categories available in a given city—are instrumental in analysing nestedness patterns within a bipartite network of cities and retail categories (Almeida-Neto et al., 2008; Atmar & Patterson, 1993). We test for the degree of nestedness of the categorical composition of retail between the different cities. A system is said to be nested when the elements found in units of analysis with less elements are likely to be found in units of analysis with more elements. The degree to which this happens is referred to as a system's "temperature" (Atmar & Patterson, 1993), in which a temperature close to 100° indicates randomness (no nested pattern, categories found in smaller subsets are not especially likely to be present in larger subsets), and a temperature close to 0° indicates complete nestedness (all categories found in less rich subsets will necessarily be present in richer ones). Almeida-Neto et al. (2008) propose a new nestedness metric based on overlap and decreasing fill (NODF $\in [0, 1]$). It considers both the degree of nestedness and the decay of diversity over subsets, yielding a value for 0 when the matrix exhibits a checkerboard pattern (thus, not random, but uniformly distributed), and 1 when it is fully nested. To measure the NODF, we create a binary matrix of presence of categories, in which cities, as rows, have the value of 1 if a category, as columns, is present, and 0 otherwise. To avoid noise and inconsistent categorisations, we include only categories if their number of shops in a city is larger than 1. We order this matrix's rows by diversity (from low to high) and its columns by ubiquity (number of cities in which the categories are present). Finally, we measure its nestedness akin to Almeida-Neto et al. (2008) using tsakim (2017, October 18/2024) $python$ package.

Another metric of diversity, in addition to richness, is Shannon's entropy. Shannon entropy has often been used at the system level, to characterise the distribution of a certain phenomenon across cities, including city size. It has also been used to describe the diversity of a certain phenomenon within a regional unit, such as land-use, sprawl, or product diversity, or to describe the economic composition of a city. However, to the best of our knowledge, the comparison of this within-entropy measurement across a system of cities, and how it scales with city sizes has not been considered. The retail entropy H of a city c is defined as:

$$H_c = -\sum_{k}^{K_c} p_{k,c} \log_2 p_{k,c} \tag{1.1}$$

where $p_{k,c}$ is the proportion of the total shops in city c belonging to category k, $p_{k,c} \in [0,1]$ and K_c is the richness, or the total number of categories per city.

At last, we calculate descriptive statistics for the metrics calculated so far. As a result, it is presented as a table with the summary statistics for the collected data and calculated metrics, namely: population, area, number of shops, total number of categories, diversity of categories, per city.

1.2.3 Scaling

Agglomeration effects and scaling laws are fundamental concepts in urban studies that describe how different phenomena scale with population and city size. If larger cities exhibit disproportionally higher (lower) concentrations of a certain feature, it is said such phenomenon scales super- (sub-)linearly with population. We follow Ribeiro and Rybski (2023) to define three different types of scaling. Firstly, the overall scaling of retail, by considering the total number of shops per city as a function of population per city N_c (estimating both the intercept J_0 and the scaling coefficient β - Equation 1.2. Secondly, we define the scaling of heterogeneity, by measuring how the richness varies with city size (estimating both the intercept K_0 and the scaling coefficient γ - Equation 1.3). Lastly, we propose to estimate category-specific scaling laws, by considering the number of shops per city per category as a function of city population (estimating, again, the category-specific intercept $J_{k,0}$ and the scaling coefficient β_k per category - Equation 1.4).

$$J_c = J_0 N_c^{\beta} \tag{1.2}$$

$$K_c = K_0 N_c^{\gamma} \tag{1.3}$$

$$J_{c,k} = J_{k,0} N_c^{\beta_k} \tag{1.4}$$

1.3 Results

We begin this results section by presenting the descriptive statistics of the data collected and processed. This includes maps of total population and total number of shops, followed by a summary statistics table with the collected data and processed metrics. Secondly, we present the results of the nestedness analysis, by plotting the nestedness binary matrix, with cities ordered by richness, and categories ordered by ubiquity, together with the result for the temperature measure. Finally, we present the coefficients found for the scaling of total number of shops, the richness of shop categories, and entropy, as well as the category-specific scaling coefficients, together with a robustness check for different city sizes.

1.3.1 Descriptive statistics

This section presents the descriptive statistics of the datasets collected and processed. We start by plotting two maps (Figure 1.1 and Figure 1.2) showing, respectively, the total population per city and the total number of shops per city.

Figure 1.1: Total population per FUA.

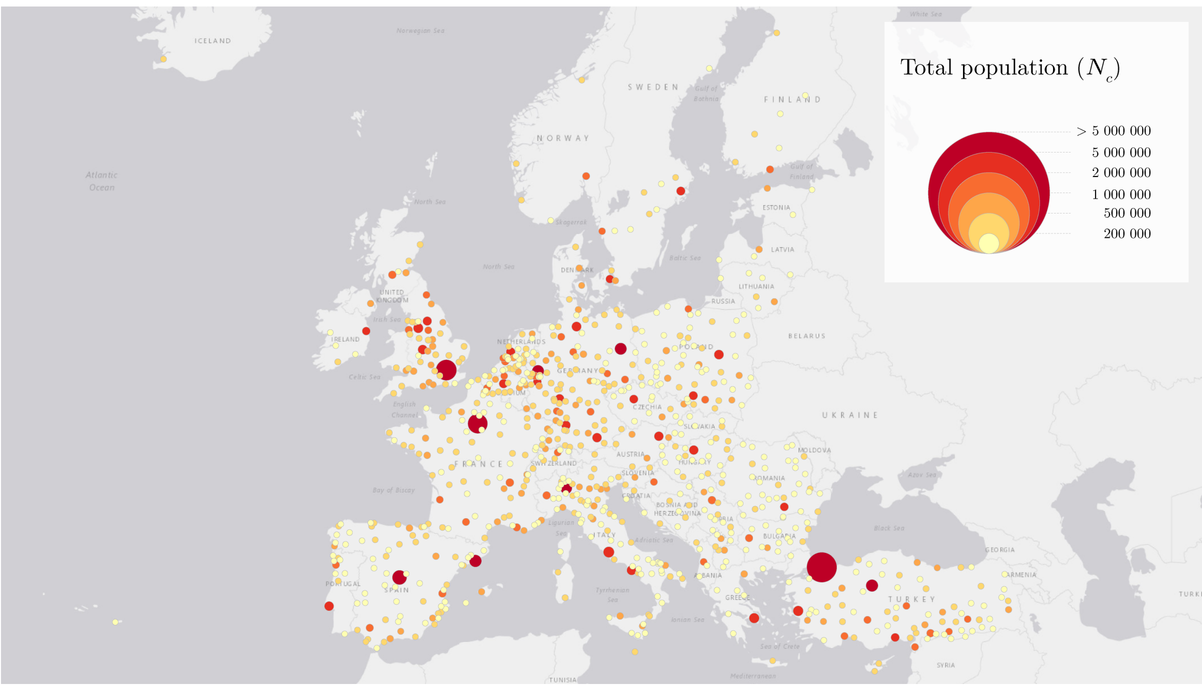
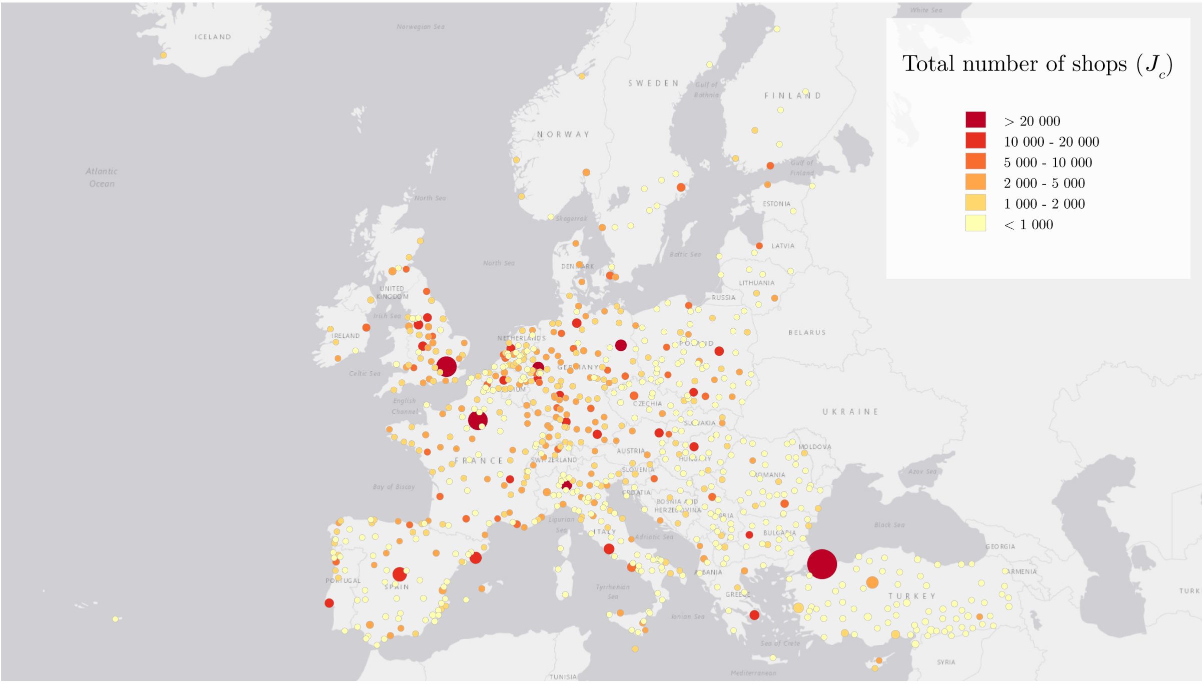


Figure 1.2: Total number of shops per FUA. Circle sizes are the same as Figure 1.1, representing population.



We, then, move to presenting a table with the summary statistics of the dataset. We can see in Table 1.2 that the count of observations for all calculations is equal to the number of cities included in the analysis (782). The cities have an average population of around 514 000, varying from just below 50 000 to a maximum of 20 891 069 (Istanbul). Then we describe the variation in the area of the FUAs in km^2. Next, we have the descriptives of the total number of shops, followed by the total number of categories (richness), and, lastly, the metric used for diversity (entropy). We include in Appendix 1.A a test of internal coherence and completeness of the FUA definitions, which includes a plot for the Zipf-law of population distribution (coefficient equal to -1.01), and the fundamental allometry (scaling of area with population - coefficient equal to $2/3$). This is an indication that the boundary definitions of the FUAs and their selection, although external to this research, are internally coherent in representing a well-behaved system of cities, since these coefficients are in-line with other studies found in literature (Cottineau, 2017; Ribeiro & Rybski, 2023).

1.3.2 Nestedness

When plotting the nestedness binary matrix, with cities as rows ordered by diversity, and categories as columns ordered by ubiquity, we visually detect some degree of nestedness. In a counterfactual scenario of a fully nested matrix ($NODF = 1$),

Table 1.2: Summary table for acquired and processed data.

Variable	Count	Mean	Std	Min	Median	Max
Population	782	514 000.56	1 193 799.11	49 973	227 145	20 891 069
Area (km^2)	782	1619.99	1765.02	18.09	1157.89	17 484.36
Shops (J)	782	1970.58	4084.28	1.00	846.00	56 720.00
Richness (K)	782	106.67	61.51	1.00	102.00	620.00
Entropy (H)	782	5.10	0.88	0.00	5.42	6.03

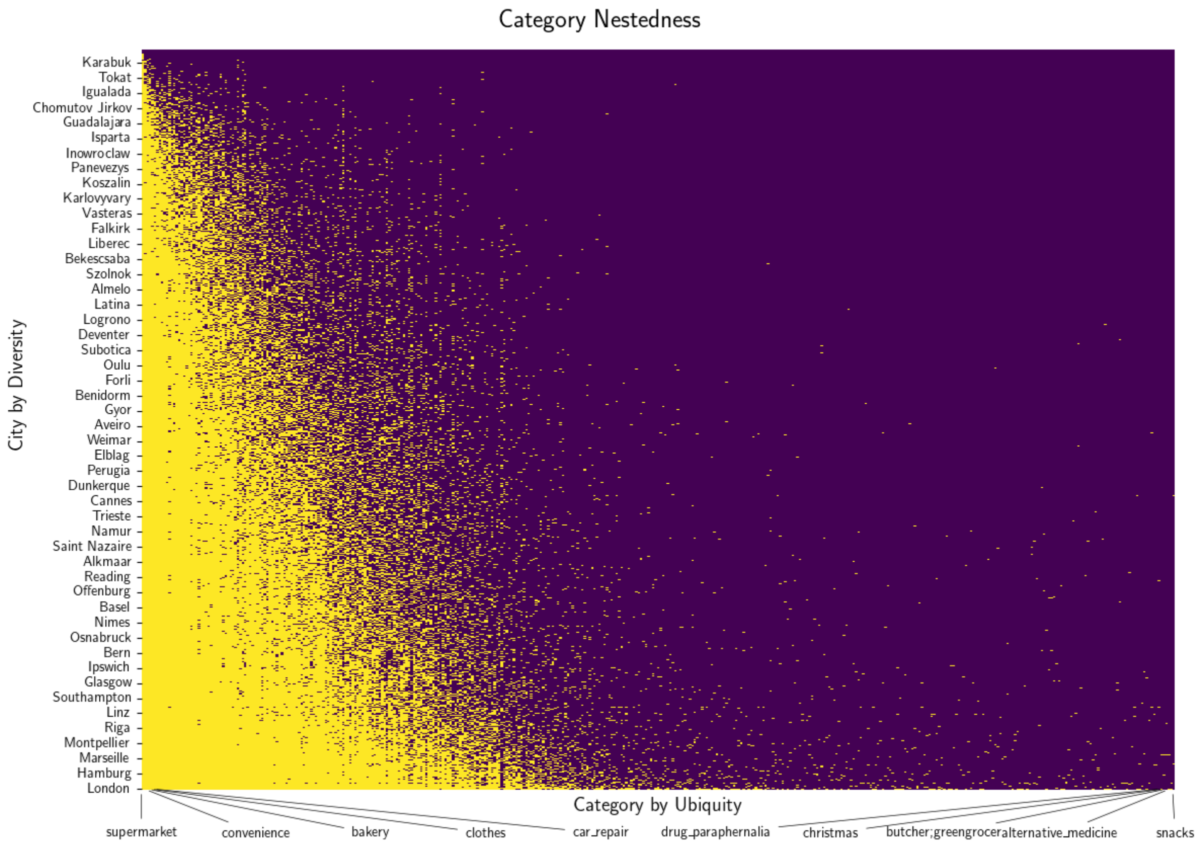
where every shop category found in less diverse cities is necessarily found also in cities with higher diversity, the yellow block in Figure 1.3 would be a solid block aligned to the bottom-left corner of the figure. In the opposite counterfactual in which no degree of nestedness is observed ($NODF = 0$) the observed pattern would generate no solid block at all, and yellow points equally scattered throughout the plot. The most ubiquitous categories of retail detected are supermarkets, convenience shops, bakeries, clothing shops, and car repair shops; while the less ubiquitous categories are snacks, alternative medicine, butcher & green grocers, Christmas and drug paraphernalia. The calculated value for $NODF$ is equal to 0.78, indicating a relatively high degree of nestedness in the categories of retail between the cities analysed.

1.3.3 Scaling

We tested the scaling of the total number of shops, of total shop categories (richness) and entropy of retail with population size for the analysed cities. While shop quantity follows a superlinear scaling with city size ($\beta = 1.14$, $r^2 = 0.52$), meaning larger cities offer a disproportionately higher number of shops, total number of categories (richness) follow a positive, but sublinear scaling ($\gamma = 0.45$, $r^2 = 0.32$). Given the left-skewed distribution of the values for entropy, we choose not to log-transform the dependent variable in this case ($Y \sim a \ln x + b$). The interpretation of the coefficients also changes in this case, no longer being an elasticity with city size, but rather solely a variation and its strength.

When moving to category-specific scaling with population, we notice a striking pattern of decreasing scaling coefficient with ubiquity. We decided to plot the estimated scaling coefficient β_k against the ubiquity ranking of each category, highlighting only the categories for which the estimation yielded a relative significance

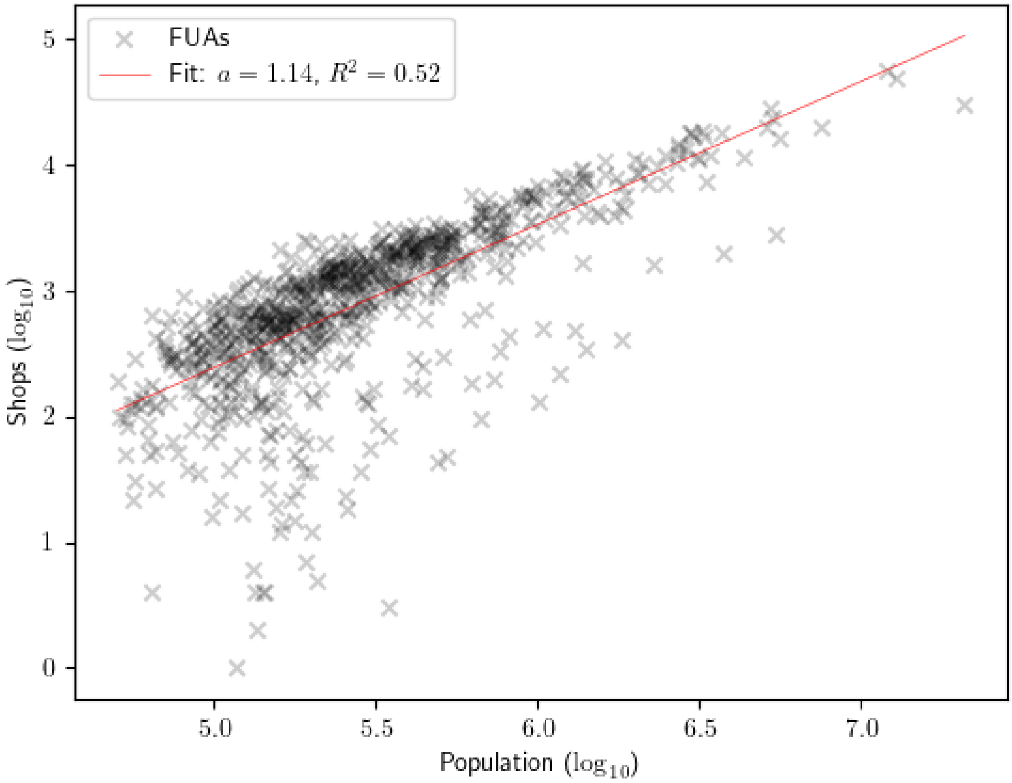
Figure 1.3: Category nestedness matrix, with rows as FUAs and columns as shop categories. Calculated degree of nestedness (NODF) equal to 0.78.



(p-value ≤ 0.01). It is clear from Figure 1.7 that the categories for which scaling is superlinear are only the most ubiquitous ones. Noticeably, scaling moves into sub-linearity, or even negative values, as we move towards rarer goods. As the relation is clearly linear, we decided to estimate yet another fitting line, relating category ubiquity u_k to scaling coefficient β_k, also excluding non-significant categories after fitting. The significant-only subset still consists of 172 categories, described in detail in Appendix 1.C. They sum 1 522 104 total shops, or 98.8% of the initial total, therefore still very representative. The fitted line indicates that an omnipresent category ($u_k = 1$) would have only a slight superlinearity, of 1.05, still below the detected overall scaling of 1.14. This is a strong indication that the overall superlinearity detected is, necessarily, *inter-sectoral*, coming together when considering an agglomeration of categories.

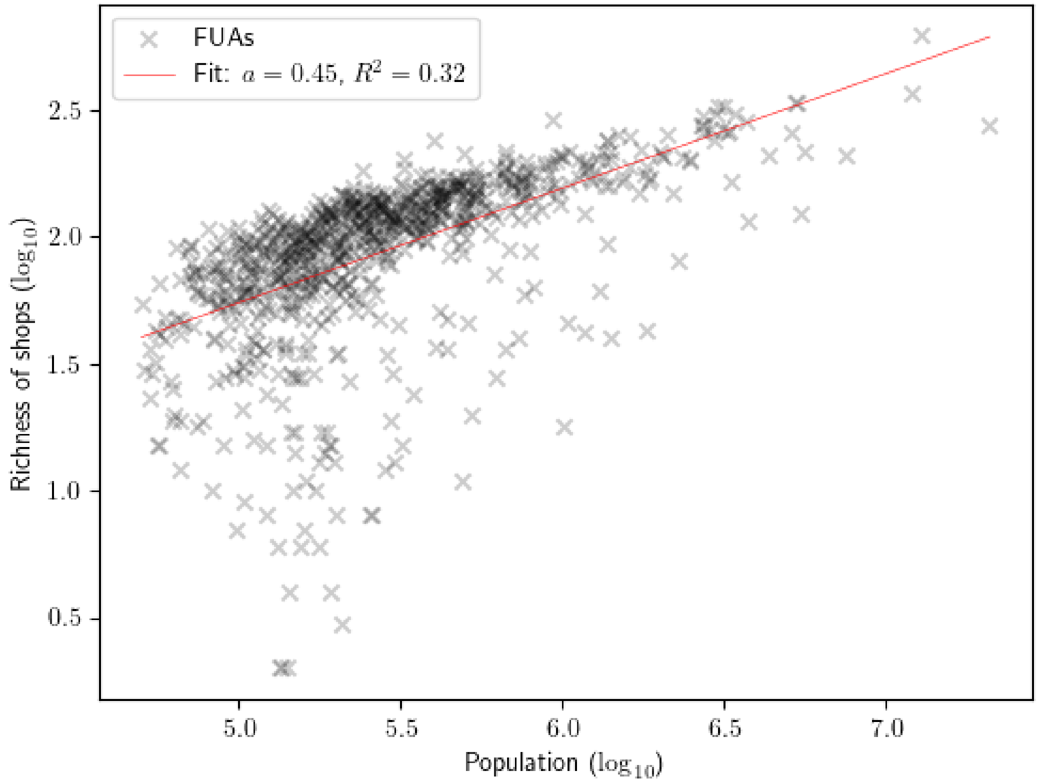
To confirm the inter-sectoral superlinearity detected previously, we decided to add one additional plot, of a cumulative scaling of categories (Figure 1.9). By gradually aggregating the number of shops from most to least ubiquitous, and estimating the

Figure 1.4: Scaling of total number of shops with population.



scaling exponent each time, we see the overall scaling converging to the previously detected one of 1.14. We see that the cumulative scaling exponent, overall, keeps on rising even when the included category has a less-than-one exponent separately. Further studies are encouraged to dive into the building blocks of this observation, perhaps making use of theoretical models to understand how this is possible. However, we dare speculate that this is due to a combined effect of diversification and specialisation. That is, the scaling of each category is added to the scaling of richness overall - that of having additional categories. In other words, the overall scaling is built both from a positive (yet less than 1) scaling of specialised categories and a positive (yet less than 1) scaling of the total number of categories. Despite the more common (ubiquitous) goods being the main force of this superlinearity, when combining the sublinearities of the rarer goods, a superlinearity is still built.

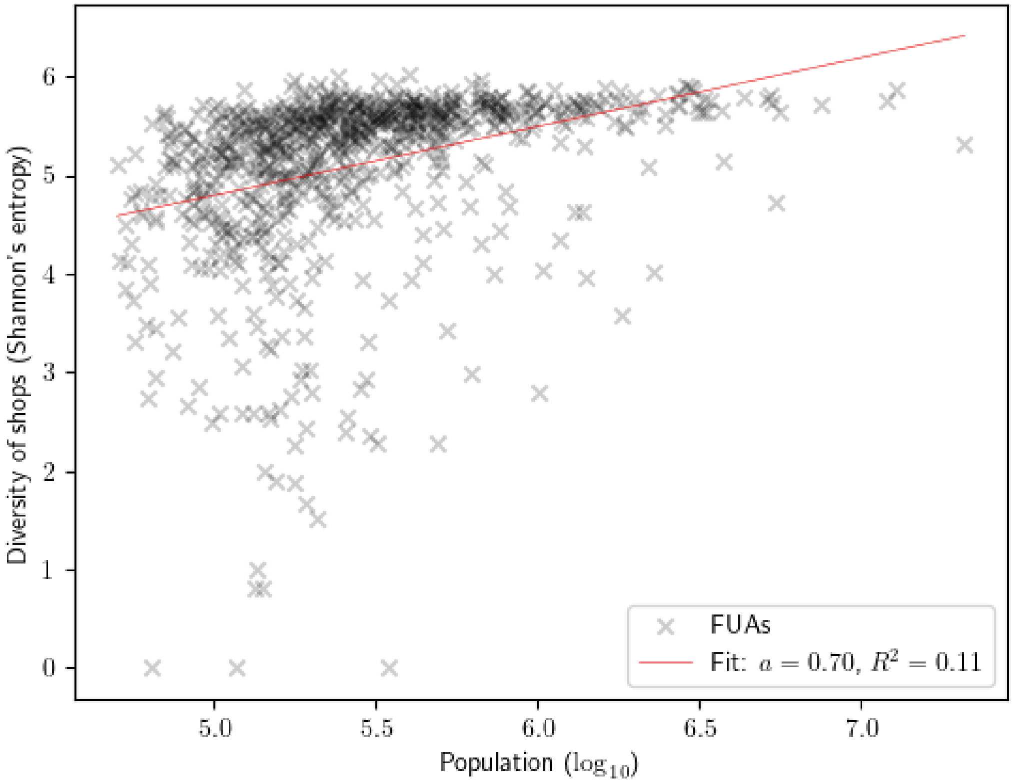
Figure 1.5: Scaling of total number of shop categories (richness) with population.



1.3.4 Robustness

As a first robustness check, we subdivide the dataset by three categories of city size (small, medium and large cities) and recalculate the overall scaling coefficient. We find that superlinear scaling of total number of shops is found in all three subsets, although with a decreasing coefficient and an increasing goodness-of-fit as we move from small to large cities. This confirms the trend observed in Subsection 1.3.3, in which total shops scale superlinearly with city size, but has two main implications. Firstly, the higher coefficients of scaling for smaller cities may indicate that either 1) a stronger agglomeration effect is present for smaller cities, with critical masses of consumers being formed as cities grow, allowing for a disproportionately larger number of shops; or 2) agglomeration effects might be capturing the amount of contributions to OSM, as a crowd-sourced platform. Secondly, the increased goodness-of-fit for larger cities, while still maintaining a superlinear coefficient, indicates that the agglomeration effects might be lower, but they are definitely more significant. The fact that larger cities, expectedly with more contributors in OSM,

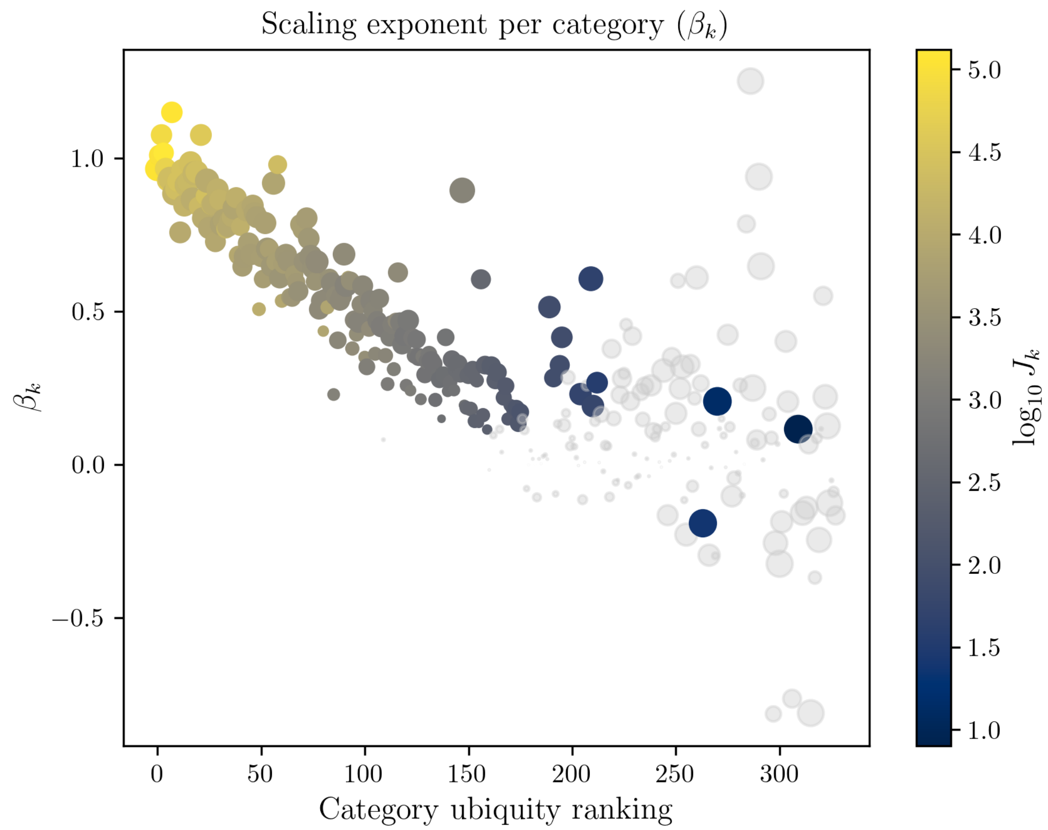
Figure 1.6: Scaling of shop category diversity (Shannon's entropy) with population.



still maintain a larger than 1 coefficient indicates that the superlinear relationship is robust, although it might be weaker than detected previously.

As a second robustness check, we colour-code the scatter plot according to each city's coordinates, detecting whether regional or country-specific patterns are determinants of outliers. The detailed definition of this and the colour-coded scatter plots are in Appendix 1.B. We highlight that the South-East corner of the colour-coding coincides with most outliers in the lower tail of the population distribution. These are, namely, Turkish cities. To control for country-specific idiosyncrasies, we estimate again the coefficients by fitting a linear mixed-effects model, grouping observations by country. The detected slope is still positive and superlinear ($\beta = 1.131$, $p = 0.024$, 95%CI $\in [1.083, 1.178]$).

Figure 1.7: Scaling of category-specific number of shops with population, ranked by category ubiquity. Greyed-out categories are the ones for which the fit did not yield a significant estimation (p-value ≥ 0.01). Colour bar represents the log of total number of shops per category.

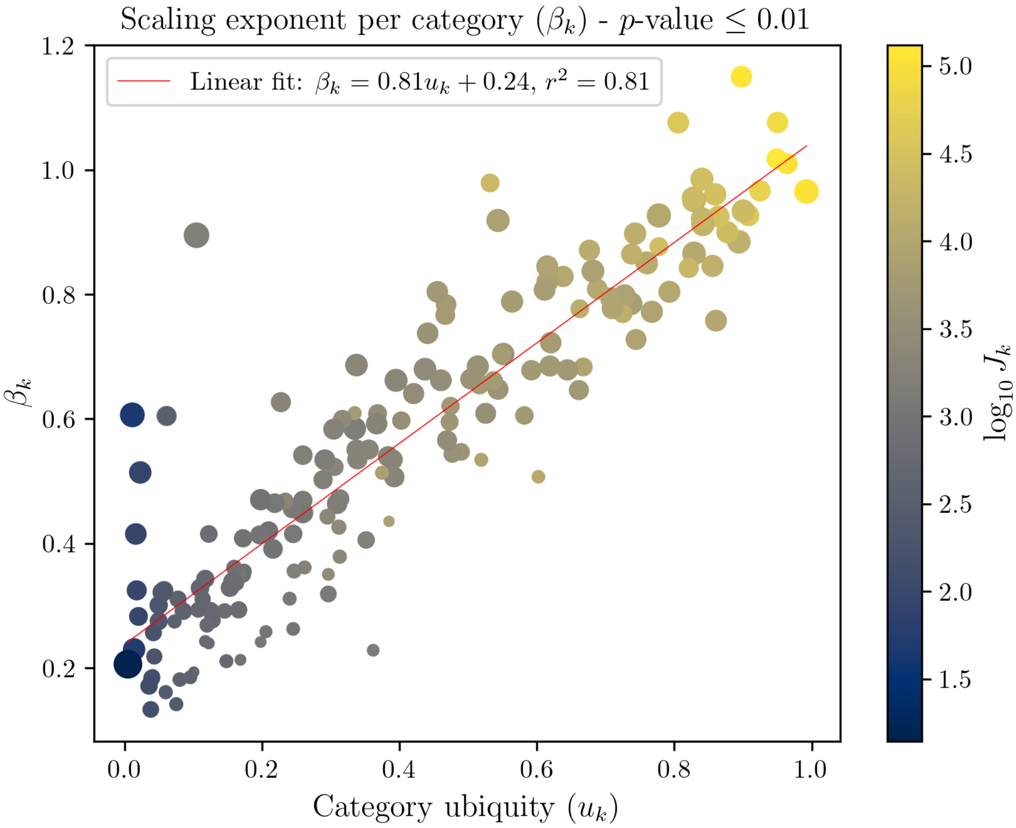


1.4 Conclusion

This study provides a detailed examination of the distribution and composition of retail establishments across 782 European cities, highlighting how retail scales with city size. We find that retail outlets, in quantity, richness and entropy, scale positively with population, meaning that larger cities not only offer more shops but also a broader variety of retail categories. Quantity, however, follows a superlinear pattern, whilst richness follows a sublinear pattern. The strength of how retail quantity scales with city size, however, also varies across city sizes: smaller cities show a stronger agglomeration effect; larger cities exhibit a more stable elasticity, closer to unitary. This confirms the role of agglomeration in shaping retail landscapes across urban systems.

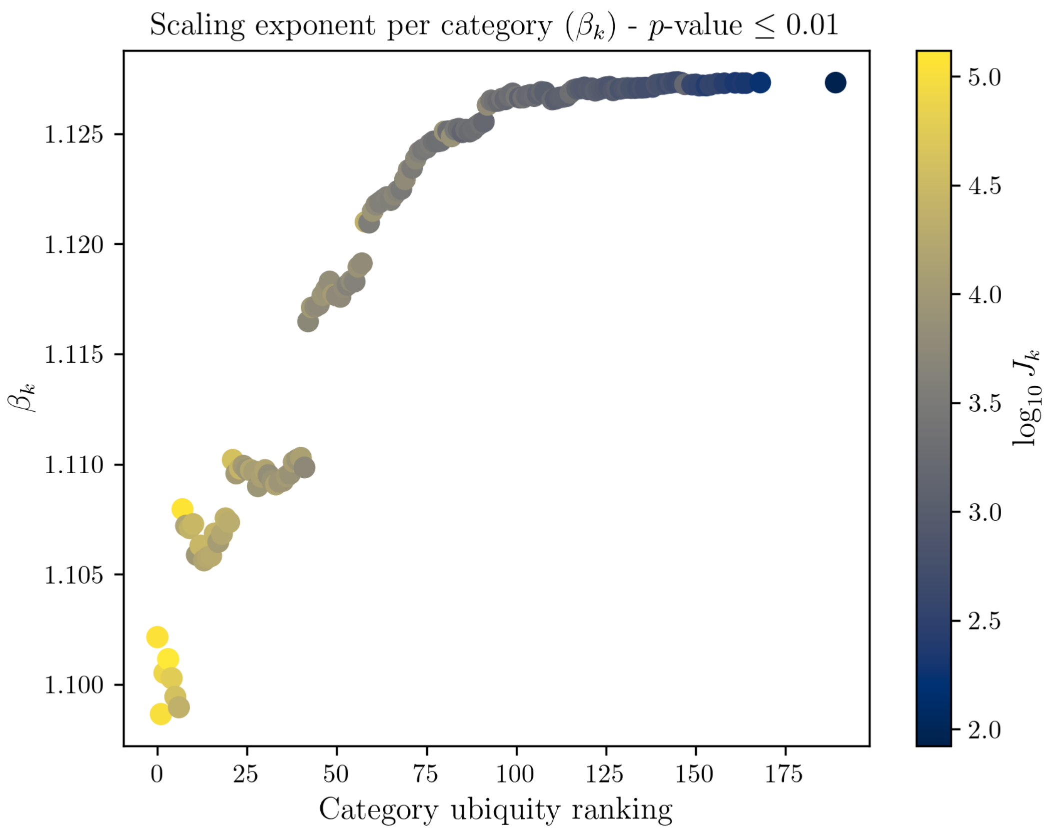
Our analysis also reveals a nested structure in the distribution of retail categories across cities, where the shop categories found in smaller cities are also present

Figure 1.8: Scaling of category-specific number of shops with population, plotted against ubiquity values after excluding non-significant categories. Colour bar represents the log of total number of shops per category.



in larger ones, but larger cities offer additional specialised (or less ubiquitous) categories. This suggests that larger cities serve as central nodes in the urban system, offering a higher degree of specialisation while maintaining common retail services found in smaller urban centres. This reenacts the Central Place Theory (Christaller, 1966), that claims that more specialised products and services tend to be present only in more central places. The high degree of nestedness observed (NODF = 0.78) underscores the hierarchical nature of retail composition, with ubiquitous categories like supermarkets and bakeries forming the core of city retail landscapes, while more niche offerings are reserved for larger urban areas. Despite nestedness being particularly relevant to the analysis of complex systems (or, more specifically, of bipartite networks such as category-spatial unit matrices), and although retail markets have been recognised as complex systems in the literature (Pennacchioli et al., 2014), our study is the first to our knowledge to estimate the degree of nestedness of retail within a system of cities. This analysis provides, then, a

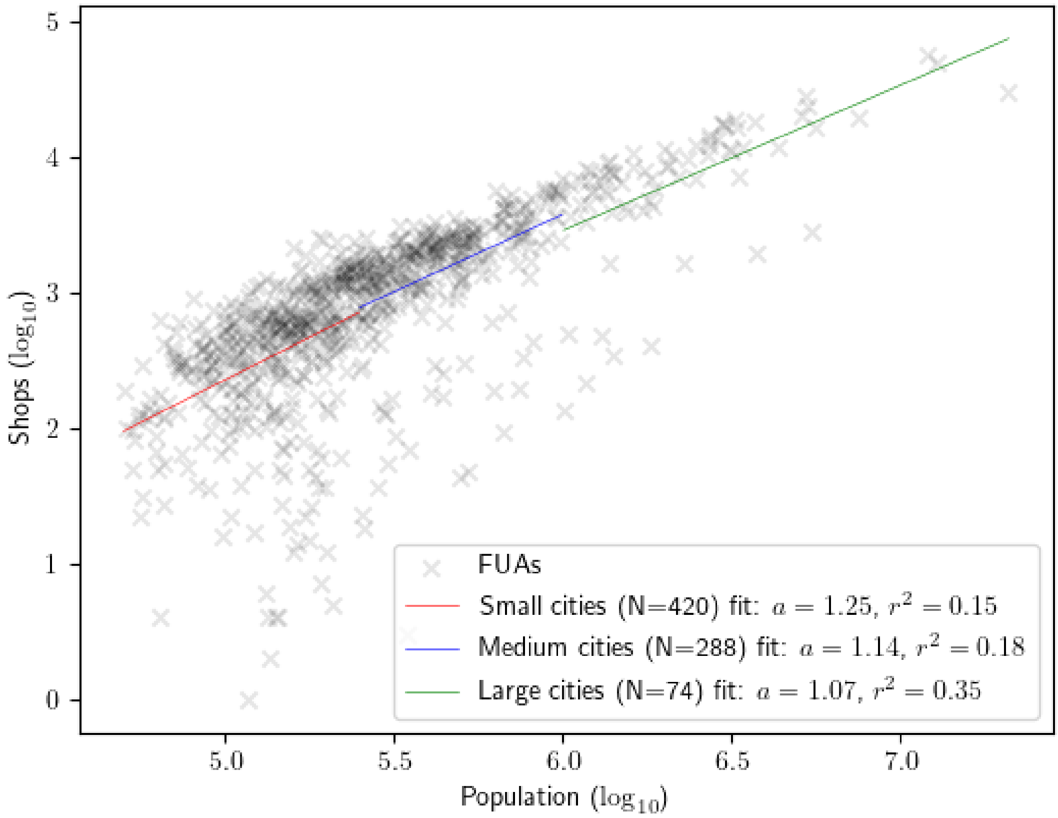
Figure 1.9: Cumulative scaling of shop categories against ubiquity ranking. Colour bar represents the log of total number of shops per category.



theoretical baseline for future research, allowing comparisons across different urban systems and encouraging the comparison between different contexts and different metrics to further explore retail hierarchies and dynamics. We also highlight the relevance of using metrics and concepts consolidated by biological sciences, or particularly ecology, in studies of the intra-urban ecosystems across systems of cities.

Our analysis reveals that the scaling behaviour of retail categories in European cities is significantly influenced by the ubiquity of goods. Specifically, we observe a decreasing scaling coefficient β_k as we move from the most ubiquitous to rarer categories, with only the most common goods exhibiting superlinear scaling. This pattern indicates that the overall superlinearity detected in retail scaling is primarily an inter-sectoral phenomenon arising from the aggregation of multiple categories. By cumulatively aggregating categories from least to most ubiquitous, we find that the cumulative scaling exponent converges to the overall value of 1.14, even when individual categories scale sublinearly. This suggests that both diversification (the

Figure 1.10: Robustness check for scaling of total number of shops with population, by different city-size categories.



addition of new categories) and specialization (increasing quantities of existing goods) contribute to the observed overall scaling behaviour. This finding directly addresses our third research question by demonstrating that the scaling of retail in European cities is driven by a combined effect of diversification and specialisation, with ubiquitous goods playing a dominant role but rarer goods collectively enhancing the superlinear scaling through their aggregated presence.

The findings of this study have significant implications for urban policy and retail planning, particularly in the context of ongoing urbanisation and shifts in consumer behaviour. The scaling laws uncovered here can inform future strategies for sustainable urban development, ensuring that retail infrastructures are adapted to meet the needs of growing urban populations. By understanding how retail diversity evolves with city size, policymakers can better support both small and large cities in fostering vibrant, accessible, and diverse retail environments.

References

Almeida-Neto, M., Guimarães, P., Guimarães Jr, P. R., Loyola, R. D., & Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. *Oikos*, *117*(8), 1227–1239.

Arvidsson, M., Lovsjö, N., & Keuschnigg, M. (2023). Urban scaling laws arise from within-city inequalities. *Nature Human Behaviour*, *7*(3), 365–374.

Atmar, W., & Patterson, B. D. (1993). The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, *96*(3), 373–382.

Bettencourt, L. M. A., & Lobo, J. (2016). Urban scaling in Europe. *Journal of The Royal Society Interface*, *13*(116), 20160005.

Brown, S. (1993). Retail location theory: Evolution and evaluation. *The International Review of Retail, Distribution and Consumer Research*, *3*(2), 185–229.

Christaller, W. (1966). *Central Places in Southern Germany*. Prentice-Hall.

Cottineau, C. (2017). MetaZipf. A dynamic meta-analysis of city size distributions. *PLOS ONE*, *12*(8), e0183919.

Jacobs, J. (1970). *The economy of cities* (1st ed.). Vintage Books.

Jonhson, S., Domínguez-García, V., & Muñoz, M. A. (2013). Factors Determining Nestedness in Complex Networks. *PLoS ONE*, *8*(9), e74025.

Kim, J., Yu, D., Choi, H., Seo, D., & Jun, B. (2024, July 29). Redefining Urban Centrality: Integrating Economic Complexity Indices into Central Place Theory.

Pennacchioli, D., Coscia, M., Rinzivillo, S., Giannotti, F., & Pedreschi, D. (2014). The retail market as a complex system. *EPJ Data Science*, *3*(1), 1–27.

Pumain, D. (2011). Systems of Cities and Levels of Organisation. In P. Bourgine & A. Lesne (Eds.), *Morphogenesis: Origins of Patterns and Shapes* (pp. 225–249). Springer.

Ribeiro, F. L., & Rybski, D. (2023). Mathematical models to explain the origin of urban scaling laws. *Physics Reports*, *1012*, 1–39.

Schiavina, M., Freire, S., & MacManus, K. (2023, April). GHS-POP R2023A - GHS population grid multitemporal (1975-2030).

Tenkanen, H. (2020). *Pyrosm* [Available at: https://github.com/HTenkanen/pyrosm].

Topf, J. (2013). *Osmium command line tool* [Available at: https://osmcode.org/osmium-tool/].

tsakim. (2024, May 23). *Tsakim/nestedness*.

Verhetsel, A., Beckers, J., & Cant, J. (2022). Regional retail landscapes emerging from spatial network analysis. *Regional Studies*, *56*(11), 1829–1844.

Appendix

1.A Internal coherence and completeness of dataset

We include a test for the internal coherence and completeness of the dataset in terms of representing a well-defined system of cities. For that, we test for a Zipf-law in population data (Figure 1.A.1) and estimate the fundamental allometry, or how area scales with population (Figure 1.A.2). In a well-defined system, these are detected in literature to have coefficients of, respectively, -1 and 2/3. These are values very close to the ones found to the current datasets.

Figure 1.A.1: Zipf-law for the system of cities selected.

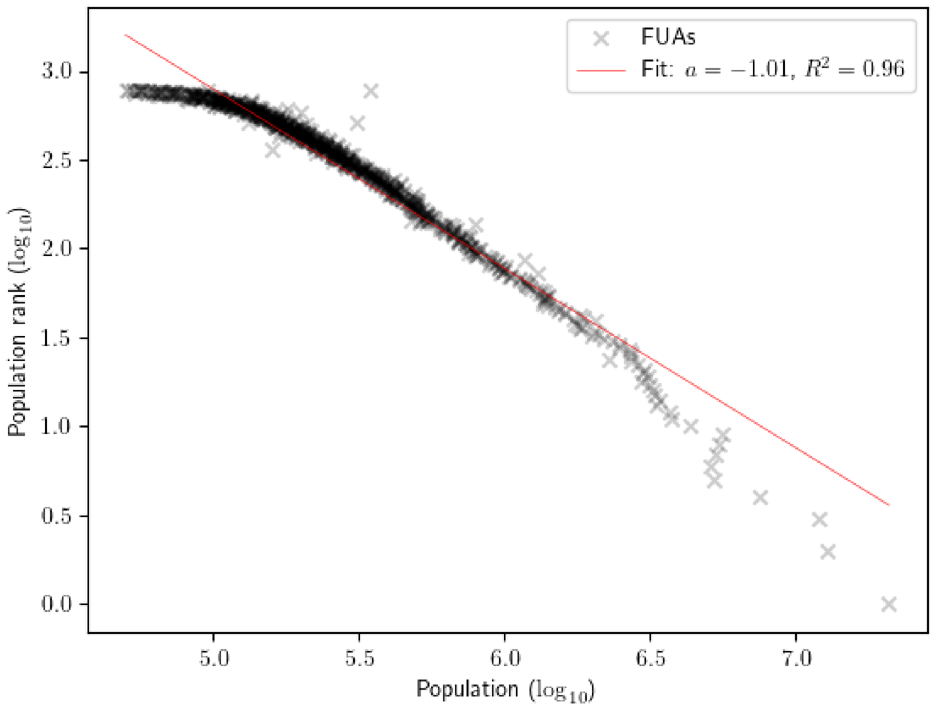
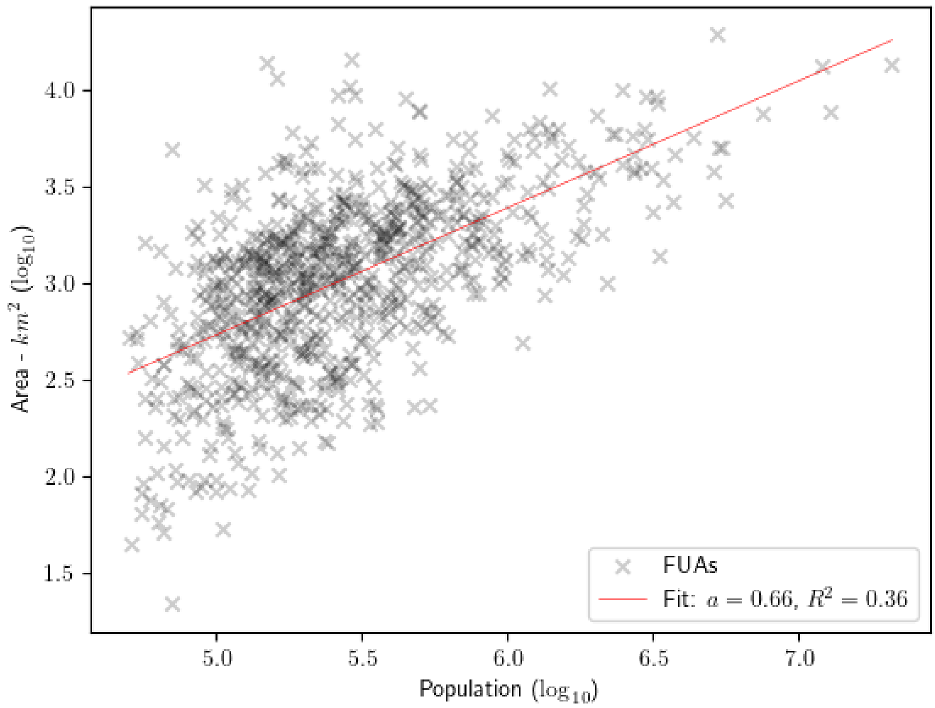


Figure 1.A.2: Fundamental allometry for the system of cities selected.



1.B Colour-coded scaling laws

As part of the robustness checks, we colour-code different cities on the scaling scatter plots based on their geographical coordinates. We that, we expect to highlight whether geographical clustering is related to the achieved metrics in each case. At first, we define the codes based on the city's locations (Figure 1.B.1). Then, we reproduce the same scaling scatter plots described in the chapter, but with the colour of each FUA according to the colour-coding defined in Figure 1.B.1. These are, respectively, the scaling of total shops with population (Figure 1.B.2), scaling of richness with population (Figure 1.B.3), and scaling of entropy with population (Figure 1.B.4).

1.C Category-specific scaling coefficients

Figure 1.B.1: Colour-code key for city locations according to their lat-lon coordinates.

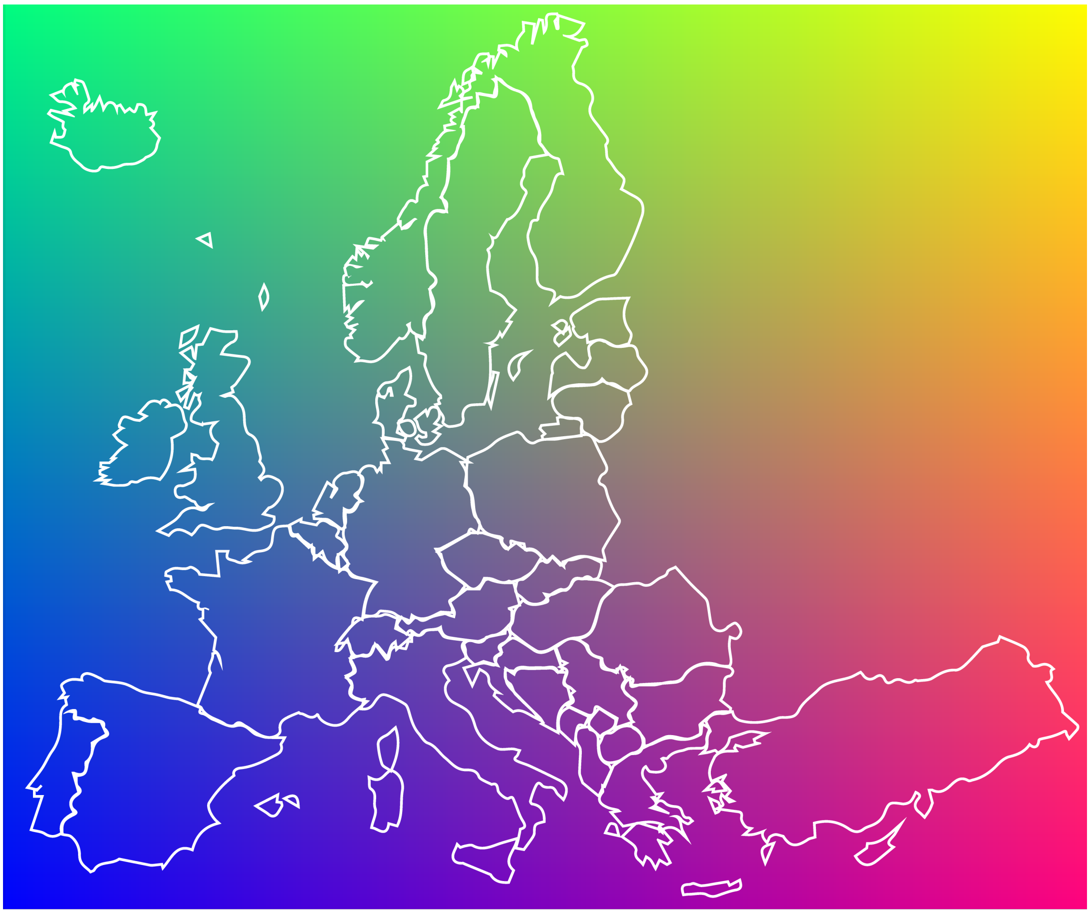


Table 1.C.1: Regression Results by Category

Category	β_k	$J_{k,0}$	p-value	r^2	J_k	u_k	Ranking
supermarket	0.965	−3.391	0.000	0.708	109 498	0.992	0
convenience	1.010	−3.809	0.000	0.495	101 657	0.964	1
bakery	1.076	−4.262	0.000	0.518	80 373	0.950	2
clothes	1.018	−3.694	0.000	0.474	131 599	0.949	3
car_repair	0.966	−3.716	0.000	0.517	58 143	0.925	4
car	0.927	−3.629	0.000	0.532	39 863	0.908	5
furniture	0.934	−3.882	0.000	0.628	23 859	0.900	6
hairdresser	1.150	−4.529	0.000	0.527	115 044	0.898	7
electronics	0.884	−3.781	0.000	0.644	15 306	0.894	8
shoes	0.900	−3.686	0.000	0.556	24 743	0.877	9
butcher	0.925	−3.750	0.000	0.546	29 770	0.864	10

Continued on next page

Table 1.C.1: Regression Results by Category (Continued)

Category	β_k	$J_{k,0}$	p-value	r^2	J_k	u_k	Ranking
mall	0.758	−3.253	0.000	0.543	9905	0.861	11
florist	0.961	−3.970	0.000	0.574	29 076	0.859	12
doityourself	0.846	−3.553	0.000	0.563	15 394	0.855	13
mobile_phone	0.912	−3.904	0.000	0.637	18 132	0.841	14
bicycle	0.922	−3.924	0.000	0.560	19 390	0.840	15
optician	0.985	−4.135	0.000	0.611	27 133	0.840	16
sports	0.866	−3.759	0.000	0.652	11 504	0.829	17
books	0.951	−4.122	0.000	0.673	16 942	0.829	18
jewelry	0.955	−4.011	0.000	0.600	23 990	0.827	19
yes	0.842	−3.505	0.000	0.452	20 397	0.821	20
beauty	1.076	−4.573	0.000	0.543	39 864	0.806	21
hardware	0.804	−3.469	0.000	0.553	10 760	0.793	22
kiosk	0.877	−3.658	0.000	0.386	26 698	0.777	23
pet	0.927	−4.156	0.000	0.690	10 603	0.777	24
computer	0.772	−3.360	0.000	0.562	8506	0.767	25
greengrocer	0.849	−3.618	0.000	0.488	16 361	0.761	26
variety_store	0.851	−3.659	0.000	0.585	12 150	0.760	27
car_parts	0.728	−3.093	0.000	0.494	8753	0.744	28
travel_agency	0.898	−3.920	0.000	0.572	13 486	0.743	29
gift	0.865	−3.679	0.000	0.500	15 256	0.738	30
toys	0.785	−3.500	0.000	0.627	6982	0.737	31
confectionery	0.798	−3.546	0.000	0.565	8199	0.728	32
chemist	0.769	−3.181	0.000	0.405	14 726	0.725	33
stationery	0.777	−3.403	0.000	0.503	9050	0.710	34
garden_centre	0.793	−3.397	0.000	0.581	9070	0.708	35
laundry	0.808	−3.511	0.000	0.489	10 940	0.688	36
cosmetics	0.838	−3.770	0.000	0.608	8090	0.682	37
alcohol	0.871	−3.814	0.000	0.505	12 620	0.676	38
pastry	0.683	−2.909	0.000	0.386	8797	0.668	39

Continued on next page

Table 1.C.1: Regression Results by Category (Continued)

Category	β_k	$J_{k,0}$	p-value	r^2	J_k	u_k	Ranking
newsagent	0.777	−3.263	0.000	0.386	13 332	0.662	40
tyres	0.646	−2.808	0.000	0.450	5576	0.661	41
department_store	0.679	−3.013	0.000	0.513	5366	0.645	42
interior_decoration	0.829	−3.604	0.000	0.511	10 383	0.638	43
houseware	0.722	−3.208	0.000	0.525	6173	0.620	44
motorcycle	0.685	−2.994	0.000	0.504	5594	0.619	45
deli	0.845	−3.779	0.000	0.550	8647	0.615	46
funeral_directors	0.820	−3.632	0.000	0.540	8238	0.615	47
copyshop	0.807	−3.674	0.000	0.539	6732	0.611	48
tobacco	0.507	−1.802	0.000	0.181	11 367	0.602	49
ticket	0.678	−2.986	0.000	0.467	5666	0.592	50
seafood	0.606	−2.553	0.000	0.372	5586	0.582	51
tattoo	0.788	−3.517	0.000	0.564	6359	0.564	52
photo	0.704	−3.195	0.000	0.573	4291	0.551	53
wine	0.704	−3.048	0.000	0.406	6488	0.549	54
perfumery	0.648	−2.867	0.000	0.500	4335	0.543	55
dry_cleaning	0.919	−4.305	0.000	0.624	7913	0.543	56
trade	0.662	−2.832	0.000	0.374	6227	0.537	57
vacant	0.979	−4.262	0.000	0.389	22 211	0.532	58
paint	0.609	−2.676	0.000	0.491	3618	0.526	59
farm	0.534	−2.013	0.000	0.190	8629	0.519	60
kitchen	0.656	−2.772	0.000	0.455	5690	0.517	61
bed	0.684	−3.062	0.000	0.559	4021	0.514	62
fabric	0.664	−2.969	0.000	0.496	4009	0.504	63
second_hand	0.547	−2.247	0.000	0.336	4407	0.490	64
wholesale	0.544	−2.314	0.000	0.332	3803	0.477	65
hearing_aids	0.621	−2.635	0.000	0.364	5121	0.474	66
medical_supply	0.595	−2.523	0.000	0.347	4547	0.473	67
outdoor	0.566	−2.471	0.000	0.445	3021	0.469	68

Continued on next page

Table 1.C.1: Regression Results by Category (Continued)

Category	β_k	$J_{k,0}$	p-value	r^2	J_k	u_k	Ranking
tailor	0.784	−3.567	0.000	0.476	6072	0.468	69
massage	0.767	−3.460	0.000	0.435	6081	0.467	70
coffee	0.662	−3.028	0.000	0.538	3297	0.460	71
art	0.804	−3.684	0.000	0.531	5399	0.455	72
antiques	0.738	−3.374	0.000	0.526	4183	0.441	73
musical_instrument	0.680	−3.164	0.000	0.594	2834	0.437	74
bag	0.641	−2.938	0.000	0.506	2992	0.421	75
e-cigarette	0.597	−2.626	0.000	0.366	3504	0.403	76
music	0.662	−3.124	0.000	0.609	2285	0.395	77
baby_goods	0.507	−2.261	0.000	0.468	2031	0.393	78
bathroom_furnishing	0.535	−2.421	0.000	0.471	2109	0.390	79
beverages	0.435	−1.497	0.000	0.114	7691	0.385	80
electrical	0.540	−2.448	0.000	0.475	2052	0.384	81
bookmaker	0.513	−1.913	0.000	0.192	7776	0.375	82
craft	0.608	−2.802	0.000	0.390	3128	0.368	83
hifi	0.592	−2.754	0.000	0.519	1952	0.367	84
agrarian	0.228	−0.732	0.000	0.151	1397	0.362	85
pet_grooming	0.551	−2.458	0.000	0.449	2219	0.355	86
general	0.406	−1.705	0.000	0.327	1753	0.352	87
fashion_accessories	0.535	−2.419	0.000	0.459	2005	0.339	88
tea	0.551	−2.517	0.000	0.449	1959	0.338	89
locksmith	0.686	−3.280	0.000	0.587	2273	0.338	90
erotic	0.584	−2.809	0.000	0.590	1506	0.335	91
charity	0.609	−2.620	0.000	0.191	6231	0.335	92
carpet	0.599	−2.722	0.000	0.385	2751	0.317	93
cheese	0.379	−1.464	0.000	0.201	2339	0.313	94
sewing	0.472	−2.063	0.000	0.402	1728	0.313	95
chocolate	0.426	−1.704	0.000	0.234	2629	0.312	96
appliance	0.464	−2.070	0.000	0.472	1536	0.309	97

Continued on next page

Table 1.C.1: Regression Results by Category (Continued)

Category	β_k	$J_{k,0}$	p-value	r^2	J_k	u_k	Ranking
health_food	0.523	−2.322	0.000	0.357	2099	0.306	98
watches	0.583	−2.724	0.000	0.488	1869	0.304	99
lottery	0.350	−1.216	0.000	0.155	2774	0.297	100
fishing	0.319	−1.278	0.000	0.293	1259	0.297	101
pawnbroker	0.443	−1.819	0.000	0.264	2313	0.295	102
lighting	0.534	−2.473	0.000	0.492	1551	0.292	103
curtain	0.502	−2.283	0.000	0.403	1668	0.289	104
boutique	0.361	−1.357	0.000	0.190	1873	0.262	105
frame	0.469	−2.139	0.000	0.411	1279	0.260	106
storage_rental	0.542	−2.469	0.000	0.421	1592	0.260	107
video_games	0.449	−2.071	0.000	0.490	1094	0.260	108
herbalist	0.356	−1.325	0.000	0.228	1692	0.247	110
fireplace	0.263	−0.943	0.000	0.182	1000	0.246	111
tiles	0.415	−1.821	0.000	0.369	1156	0.246	112
leather	0.456	−2.024	0.000	0.407	1267	0.244	113
rental	0.311	−1.164	0.000	0.191	1205	0.240	114
estate_agent	0.468	−1.951	0.000	0.292	2088	0.234	115
frozen_food	0.627	−2.804	0.000	0.450	2142	0.228	116
shoe_repair	0.465	−2.113	0.000	0.426	1138	0.219	117
window_blind	0.391	−1.717	0.000	0.425	913	0.216	118
games	0.420	−1.938	0.000	0.409	844	0.210	119
no	0.258	−0.878	0.000	0.162	981	0.206	120
party	0.470	−2.212	0.000	0.514	901	0.198	121
telecommunication	0.242	−0.771	0.000	0.125	1126	0.198	122
nutrition_supplements	0.414	−1.837	0.000	0.397	980	0.197	123
flooring	0.355	−1.506	0.000	0.249	891	0.174	124
motorcycle_repair	0.408	−1.848	0.000	0.362	836	0.173	125
hairdresser_supply	0.351	−1.526	0.000	0.348	721	0.171	126
dairy	0.213	−0.669	0.000	0.125	859	0.169	127

Continued on next page

Table 1.C.1: Regression Results by Category (Continued)

Category	β_k	$J_{k,0}$	p-value	r^2	J_k	u_k	Ranking
cannabis	0.352	−1.380	0.000	0.223	1032	0.168	128
radiotechnics	0.293	−1.218	0.000	0.315	630	0.166	129
glaziery	0.337	−1.420	0.000	0.282	782	0.162	130
doors	0.361	−1.512	0.000	0.272	860	0.160	131
video	0.339	−1.466	0.000	0.386	650	0.157	132
weapons	0.329	−1.413	0.000	0.369	641	0.153	133
caravan	0.211	−0.708	0.000	0.193	544	0.148	134
tool_hire	0.292	−1.144	0.000	0.244	665	0.146	135
pottery	0.277	−1.115	0.000	0.300	514	0.128	136
repair	0.292	−1.228	0.000	0.314	523	0.125	138
grocery	0.415	−1.873	0.000	0.331	686	0.123	139
swimming_pool	0.240	−0.814	0.000	0.127	597	0.123	140
household_linen	0.269	−0.987	0.000	0.229	579	0.120	141
model	0.343	−1.552	0.000	0.359	488	0.118	142
ice_cream	0.243	−0.854	0.000	0.144	612	0.118	143
collector	0.311	−1.335	0.000	0.273	476	0.114	144
printer_ink	0.328	−1.475	0.000	0.379	449	0.110	145
food	0.293	−1.232	0.000	0.260	480	0.107	146
convenience;gas	0.895	−3.821	0.000	0.778	1568	0.105	147
money_lender	0.194	−0.518	0.000	0.110	534	0.101	148
scuba_diving	0.185	−0.608	0.000	0.188	378	0.096	149
security	0.291	−1.233	0.000	0.324	376	0.086	150
religion	0.181	−0.568	0.000	0.210	351	0.081	151
printing	0.311	−1.376	0.000	0.307	352	0.078	152
fuel	0.142	−0.423	0.000	0.195	246	0.075	153
spices	0.275	−1.212	0.000	0.198	347	0.073	154
catalogue	0.605	−2.907	0.000	0.447	359	0.061	156
hunting	0.161	−0.510	0.000	0.199	258	0.060	157
camera	0.325	−1.516	0.000	0.381	263	0.058	158

Continued on next page

Table 1.C.1: Regression Results by Category (Continued)

Category	β_k	$J_{k,0}$	p-value	r^2	J_k	u_k	Ranking
wool	0.322	−1.410	0.000	0.413	212	0.055	161
car;car_repair	0.275	−1.232	0.000	0.350	230	0.050	163
honey	0.301	−1.350	0.000	0.369	205	0.050	164
wigs	0.218	−0.911	0.000	0.269	184	0.043	167
gold_buyer	0.257	−1.008	0.001	0.306	175	0.042	168
pyrotechnics	0.186	−0.643	0.004	0.242	179	0.041	172
trophy	0.184	−0.753	0.001	0.300	147	0.040	173
garden_furniture	0.133	−0.413	0.002	0.291	144	0.038	174
equestrian	0.171	−0.607	0.002	0.323	126	0.036	175
fan	0.514	−2.638	0.000	0.579	84	0.023	189
plant_hire	0.283	−1.257	0.010	0.388	88	0.020	191
water	0.325	−1.514	0.009	0.449	90	0.018	194
household	0.415	−1.914	0.005	0.529	85	0.017	195
number_plate	0.230	−0.830	0.007	0.578	53	0.014	204
smartshop	0.607	−2.994	0.004	0.722	46	0.012	209
homeware	0.206	−0.862	0.001	0.997	14	0.005	270

Figure 1.B.2: Scaling of total number of shops with population (colour-coded by city coordinates).

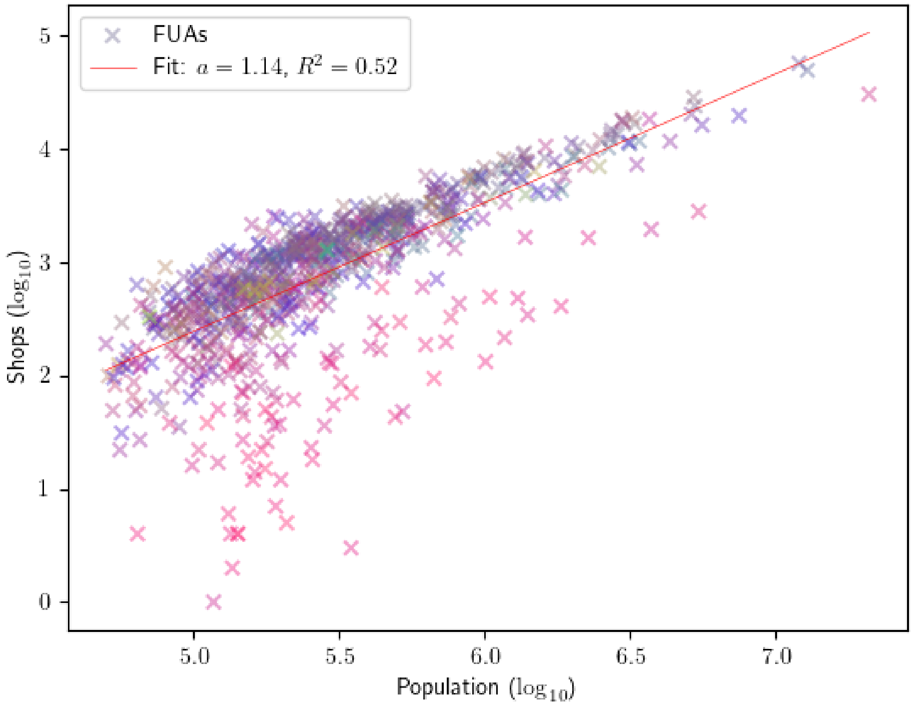


Figure 1.B.3: Scaling of total number of shop categories (richness) with population (colour-coded by city coordinates).

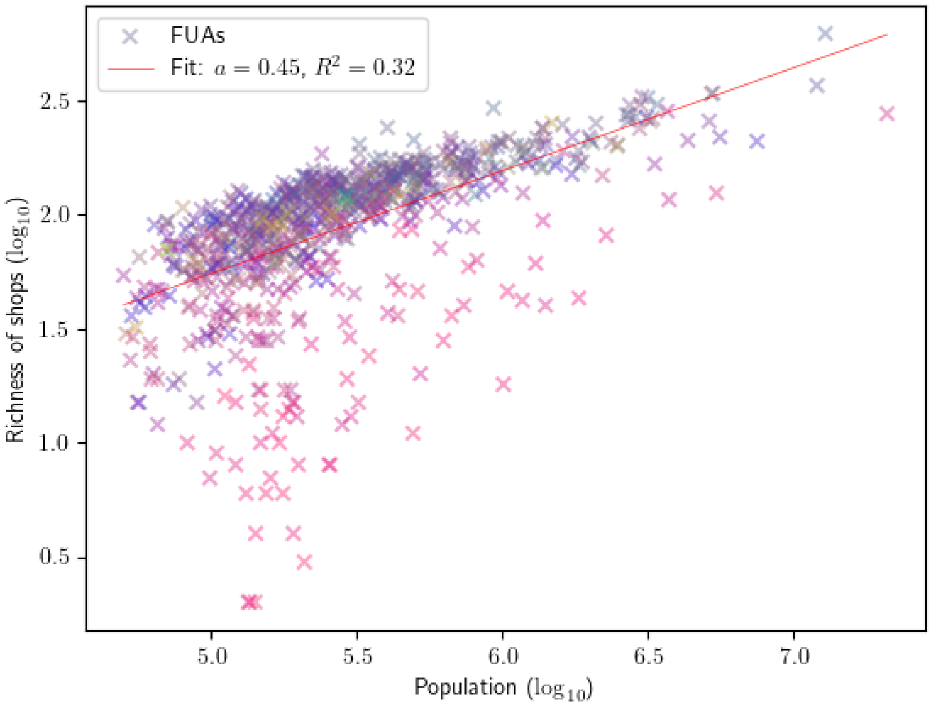
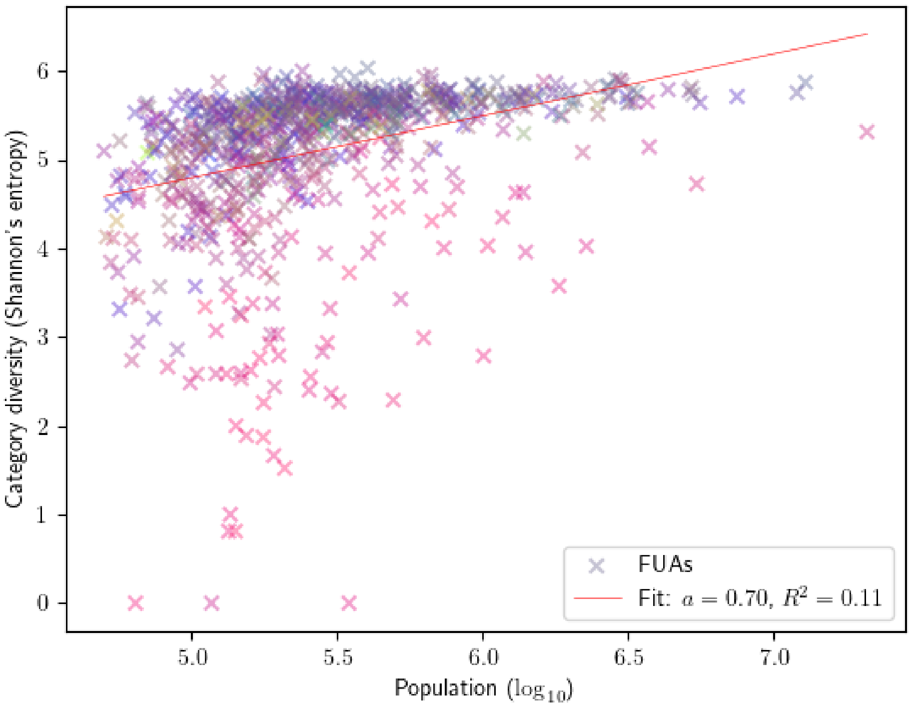


Figure 1.B.4: Scaling of shop category diversity (Shannon's entropy) with population (colour-coded by city coordinates).



Chapter 2

The *retailscape* of European cities: assessing the 15-minute accessibility of retail location with open data

Retail constitutes a key function of urban life. Shops not only connect demand to supply of both everyday and specialised goods, but they also serve as an important provider of amenities for citizens, especially with the rise of online-shopping. This study analyses retail accessibility across 782 European Functional Urban Areas (FUAs) using Open Street Map (OSM) data for retail location, and the Global Human Settlement Layer (GHSL) for population data. We analyse spatial patterns of retail distribution across cities and compare them to local population density to highlight local and regional differences in accessibility to retail. Firstly, we calculate travel times between intra-urban hexagonal grid cells and shops, accounting for space frictions using the concept of accessibility. Then, we calculate, for each FUA, a series of time-bounded accessibility metrics and a population-weighted city-level accessibility indicator, allowing for cross-city comparisons and country-level aggregations. Lastly, we propose a *retailscape* tightness measurement by comparing two time-bounded accessibilities, under the lens of the 15-minute city concept.

Findings reveal that middle-sized Mediterranean cities, especially in Spain and Greece, exhibit the highest accessibilities to commerce within shorter travel times, while larger metropolitan areas like Paris, London, and Berlin dominate longer thresholds. Finally, we highlight that regions of Spain, Southern Italy, Albania and Greece exhibit a highly tight *retailscape*, approximating them to the model advocated by the 15-minute city. This study focused on observed patterns of retail distribution, encouraging further research to investigate specific mechanisms that make these patterns emerge.

2.1 Introduction

Shopping constitutes an important dimension of urban life. Ever since medieval times, towns have had the role of facilitating trade between citizens. It is the place where producers of the surrounding farmlands have gathered to access the market of citizens in need of those products. With time, the form in which this exchange takes place has evolved: from traditional markets in the town centres to formalised shops, and eventually shopping centres that concentrate this exchange. Nevertheless, this evolution came as complementing previous forms rather than substituting: nowadays there is a complex overlay of different shopping forms in our cities. But they all exhibit a similar phenomenon, that seems to benefit sales at least to some degree: the spatial clustering of vendors. This may happen for several reasons: agglomeration of shops may share common and initial fixed costs, such as security or logistics; as a result of competition itself, as the equilibrium location for two competing firms would be the centre of their combined markets (Hotelling, 1929). Shops may also be solely following demand: people tend to cluster to be close to their daily needs, so shops cluster to be close to their potential clients.However, similarly to other agglomerating phenomena (population, jobs, etc.), we see that they do not all agglomerate in a single spot, indicating that there is an incentive for a disagglomeration as well. Dispersion forces may include limitations to firms' market catchment areas, transportation and delivery costs, or even restrictions in land-use availability. The interaction between agglomeration and dispersion is mediated by space, and its role in strengthening or weakening one or the other, e.g., transportation networks facilitate mobility, while congestion hinders it. This is a strong sign that local factors (such as the disposition of the transportation network) affect greatly the spatial patterns of shop agglomeration.

By benefiting from agglomeration economies and spatial spillovers of population concentrations, shops exhibit a superlinear scaling with city size, similar to the expected scaling of socio-economic activities with population (Ribeiro & Rybski, 2023). The specific forms and implications of this have been discussed in Chapter 1. Higher population densities expected in larger urban areas, especially in its centres (Lemoy & Caruso, 2020), may drive this superlinearity, as shops tend to require a critical mass of potential consumers to be economically viable. This, however, does not mean larger cities offer a more *accessible* retail network to its citizens. In other words, the inhabitants of larger cities do not necessarily experience this superlinearity, as the intra-city spatial patterns of shop agglomeration may differ from that of population concentration. Other local factors, such as land-use regulations and urban morphology, may influence how both shops and people distribute in space, potentially generating a mismatch between the two distributions. This brings us to our first research question: how does the spatial distribution of population within a city compare to the distribution of retail establishments?

In addition to the spatial distribution of people and shops, it is crucial to consider the infrastructure connecting them - namely, the street network. Despite possible mismatches between the spatial patterns of population and retail, they could still be arranged in such a way that shops remain easily reachable by the population via the street network. This ease of reaching opportunities and services is consolidated in literature as the concept of *accessibility* (Wachs & Kumagai, 1973). Accessibility to a wide range of urban phenomena is expected to vary within cities, as more compactly built urban tissues diminish distances as compared to sprawled suburbia, but also between cities, as larger cities extending over a larger area may concentrate amenities in specific, central places, far from population centres. The need to consider the real availability of shops to city inhabitants, beyond simple proximity or spatial coincidence, brings us to our second research question: how does the accessibility to retail vary within and between different cities?

The desirability of accessibility to retail as an essential urban amenity, or, put simply, accessibility as a positive feature cities can offer, has been put forward by different so-called normative urban theories and models. Recently, the "15-minute city" conceptual model (Moreno et al., 2021) has gained traction by proposing that compact, efficient urban spaces should offer essential services and goods, including retail, within a 15-minute walk or bike ride from residents' homes. This approach not only aims at reducing transportation-related emissions but also seeks to foster local economies and community well-being by promoting local buying. Despite proximity as a positive conotation in urban theory not being a novel concept, with examples of similar approaches found in the "neighbourhood unit" and the "compact city", the "15-minute city" helped shed a recent light on the existing conditions of cities, on whether or not they already adhere to such configuration, and to which extent. Moreover, the value of 15 minutes may seem ad-hoc, but adds to previously qualitative approaches by proposing an essentially quantifiable travel-time threshold to guide planning practises in different contexts. Although both the conceptual model and state-of-the-art research in the field consider an amalgam of services or amenities to be reached within this defined travel time, we choose to focus on the sphere of retail only. Under the light of the "15-minute city", we then propose our third research question: how can the level of "15-minuteness" be assessed for the sphere of retail for different cities?

This research, then, aims to analyse retail and population distributions within cities, assessing and comparing their accessibility to shops through the lens of the "15-minute city" model. To answer the described research questions, we select and compare 782 European cities of different sizes, across 38 countries. We use Urban Atlas' Functional Urban Areas (FUAs) to define their boundaries, extract population data from Global Human Settlement Layer (GHSL), and analyse their internal *retailscape* using shop location and street network data extracted from OpenStreetMap (OSM), and city-level hexagonal grids defined from H3. We start at a local level by calculating network travel times and quantifying proposed accessibility metrics and move towards aggregating and comparing cities at a broad geographical scale. By doing this, we expect to highlight local and regional differences crucial for urban planning and policy-making, as it has significant implications for economic development and overall quality of life in cities.

This paper is organised as follows. First, we analyse the background literature in retail location patterns, accessibility calculation, and proximity-based policy frameworks such as the "15-minute city". We then move to defining the data used in this analysis, namely the city boundary definitions, hexagon-grid definitions, population data, shop location data, and street network. This is followed by the methodological section: calculation of travel matrices, proposed accessibility metrics, strategies to aggregate accessibility metrics at the city level, and the proposed index to assess the adherence to a 15-minute city for the selected cities. Then, we describe the results of the proposed metrics and index, and, finally, the conclusion.

2.2 Background

Retail concentration, its causes and consequences, have been thoroughly studied both in geographical analyses and theoretical economics literature. From the latter, we highlight the seminal work of Hotelling (1929) claiming that the equilibrium location of two identical firms competing for a consumer market over a line would be the very centre of the line. Later, studies have further elaborated that this equilibrium is very specific, and slight changes to the settings, such as the transportation cost function, could alter that. This is the case of d'Aspremont et al. (1979), who introduced quadratic transportation costs and reached the opposite equilibrium configuration, i.e., the farthermost points of the line. With differentiated firms, Christaller (1966) proposed a conceptual model in which more specialised places would tend to locate in more central places, which was called the Central Place Theory. The idea behind it was later applied to microeconomic theoretical models and numerical experimentation (Becerra Valbuena, 2013; Fujita et al., 1999; Openshaw & Veneris, 2003; Taylor & Hoyler, 2021), highlighting the fractal configuration specialisation may generate.

Retail distribution is also claimed to vary with local and regional cultural characteristics (Cachinho, 2014). In the sphere of geographical analyses, Araldi and Fusco (2019) highlight how shopping concentration areas can take different forms and functions in a study in Southern France. Hogba and Yiran (2024) analyse the distribution of retail in sub-Saharan Africa, highlighting the importance of retail locations in city growth. Efeoglu et al. (2023) analyse retail distribution in Helsinki, focusing on the role of plot configuration for that distribution. We found further studies either focusing on retail for specific regions or on general accessibility and walkability scores for cross-city comparisons, but to our knowledge, no retail accessibility assessment has been conducted in a systematic manner across the European continent.

The concept of accessibility as the ease to reach opportunities and services (Wachs & Kumagai, 1973) is consolidated in literature. Boisjoly and El-Geneidy (2017) highlight the importance of including accessibility metrics in planning discussion and practise. It has been applied together with the adjective "cumulative", which indicates the total number of a certain destination (e.g., jobs, schools, shops) within a given travel time threshold (Pereira et al., 2019). The very choice of this threshold is often the object of analyses, as *ad-hoc* choices can alter the conclusions of assessments (Tomasiello et al., 2023). Accessibility can also be weighted by distance, considering stronger effects for closer destinations (Hidalgo et al., 2020). Normative conceptual models, such as the 15-minute city model (Moreno et al., 2021) have recently entered the discussion providing a public debate on time-based thresholds for accessibility assessments. The model proposes that a set of destinations should be reached by the citizens within a 15-minute travel window. The specific definition of these destinations varies from study to study, sometimes highlighting their composition (e.g., diversity of amenities), other times just the number of opportunities (e.g., jobs) reached.

The use of open data, more specifically OpenStreetMap for amenities, is also conducted by Olivari et al. (2023), evaluating the 15-minute city model in two cities of Italy. Broader analysis at a continental scale, as proposed by this paper, is also conducted by Bartzokas-Tsiompras and Bakogiannis (2023), highlighting German cities as best-performing and UK cities as worst-performing, although focusing on walkability in general rather than retail access. Bartzokas-Tsiompras et al. (2023) focus on accessibility to a multitude of amenity destinations at a broader scale, but again not focusing on retail as we propose. The work of Barbieri et al. (2023) focuses specifically on services, and uses a graph-based accessibility analysis like the one proposed for this paper, emphasising diversity of destination categories also under the 15-minute city lens. Under the same lens, we also highlight the works of Ulloa-Leon et al. (2023), Vale and Lopes (2023), Ferrer-Ortiz et al. (2022), all focusing on specific cities rather than cross-city analyses, and again without necessarily focusing on retail. The importance of retail in accessibility analysis is concluded by Graells-Garrido et al. (2021), finding that people tend to visit neighbourhoods with higher access to retail.

In this study, we bridge the literature of retail fabric assessment, cumulative accessibility, and the 15-minute city conceptual model. Several studies conduct spatial analyses of the retail fabric in European cities, but to our knowledge few propose a systematic cross-city and cross-country comparison at a continental scale. Others have deeply analysed cumulative accessibility indicators and indices, understanding how easily multiple urban phenomena are accessed by population, but cumulative accessibility often fails to weigh for spatial friction, which we propose to address. Finally, the normative 15-minute city conceptual model has been advocated in public debate, and studied in a quantitative manner, but rarely focusing thoroughly on the sphere of retail. In this paper, we propose to use the lens of a 15-minute city conceptual model to analyse the retail fabric of Europe, proposing accessibility indicators to compare different urban areas' internal structures.

2.3 Data definition

We utilise the dataset described in Chapter 1, with the same definition of Functional Urban Aras (FUAs), the same retail location points, and the same source for the population dataset. To conduct intra-urban analysis, we define a hexagon grid overlayed on top of each FUA to conduct the accessibility analyses. Using also OpenStreetMap (OSM), we extract the street network for each FUA to calculate the network distances used to derive the accessibility metrics. The hexagon grid is also used to resample the population data collected from Urban Atlas to conduct the analysis in the intra-urban level. For OSM data, we choose to extract data for the whole continent and later clip it by the geographical units' boundaries so that both the location of shops and street network are accounted for.

2.3.1 Hexagon grids

To conduct intra-city accessibility analyses, we choose to use H3's hexagonal gridding system, in combination with GeoPandas package (Jordahl et al., 2020), using the aggregator package H3-Pandas (Dahn, 2021). The choice of a hexagonal grid allows for reduced edge and neighbourhood effects as compared to traditional squared grids (i.e., the immediate neighbours of a cell are at an approximately equal distance to the cell's centre). Moreover, the H3 gridding system has an open-source, standardised definition to facilitate the reproducibility of the results. We choose H3's resolution of 8, which generates a grid of hexagons of, on average, approximately 530m of side length, 460m of in-radius, 0.73 km² of area (or 73 ha). The resolution was chosen so that the walking time of the distance between the centroids of two adjacent hexagons (two times the in-radius) would still be under the lowest bound of the accessibility analysis (15 minutes, following (Moreno et al., 2021), for further information see Section 2.4.3). Given a walking speed of 4km/h, this means 13.8 minutes for a straight line between two adjacent polygons' centroids. A finer resolution (≥ 9) for the gridding system could also be possible, but would exponentially increase the computational expenses. Individual hexagons are identified as the subscript i. The total number of hexagons i per city c, also a proxy for a FUA's area given the approximate equal area of each hexagon across cities, is defined as I_c. With the grid defined, the clipped and filtered data, we move on to the next subsection, describing the network-based distances.

2.3.2 Street network

As described in the section preamble, the data extracted from OSM is used both for getting the retail location points and the street network used for distance calculations. The calculation of the distances is further described in more detail in section 2.4.1. The street network is compiled within the OSM Protocolbuffer Binary Format (PBF) file for each FUA, and is used as input for calculating the network-distances using the package r^5py.

2.3.3 Population data

The last dataset used in the analysis is population data from GHSL (Schiavina et al., 2023). Similarly to Chapter 1, we take 2020's 100m-resolution raster, with which we intersect the hexagons i per city c. This is equivalent to a resampling technique from a square-grid to a hexagon-grid. The resampling provides us with a population value n per hexagon i, defined as n_i. The aggregate population per city is defined as $N_c = \sum_i^{I_c} n_i$.

2.3.4 Nomenclature summary

We summarise the nomenclature described so far in Table 2.3.1. The 782 cities C are indexed by subscript c; the total number of shops J per city c is defined as J_c and each shop is identified by subscript j; the population n per hexagon i, defined as n_i, and the aggregate population per city is defined as $N_c = \sum_i^{I_c} n_i$.

Table 2.3.1: Nomenclature summary

Element	Identifier	Aggregator	Summaries	Meaning
Hexagon	i	I	I_c	Total hexagons in city c
City	c	C	C	Total number of cities
Shop	j	J	J_c	Total shops in city c
			J_i	Shops in hexagon i
Population	n	N	N_c	Total population in city c
			n_i	Population in hexagon i

2.4 Methods

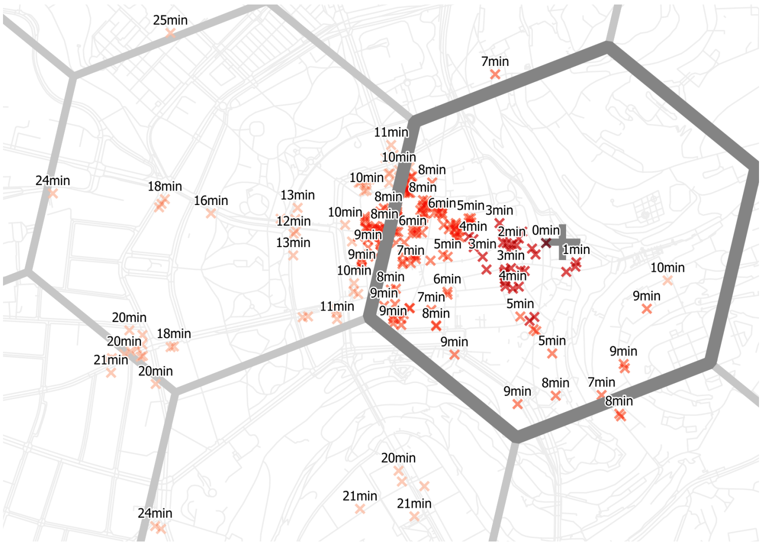
With the pre-processed data defined, we move on to describing the methods applied to them. Firstly, we choose to plot the number of shops against the population for each hexagon in all FUAs, coloured by total city population. Secondly, we propose to calculate travel matrices from all hexagons to all shops within an FUA. Thirdly, we discuss different distance-decaying functions to be applied to travel times to account for spatial friction, here called accessibility metrics (for each hexagon-firm pair), and an accessibility indicator at hexagon-level. Then, we define a series of time-bounded thresholds, with which aggregations at the hexagon-level are made. Next, we propose aggregations of these indices per city, both unweighted and weighted by population at each hexagon, in order to compare their level of accessibility to retail at the different thresholds proposed. Finally, we propose a measurement of 15-Minute Tightness and evaluate which European urban areas offer a tighter retail setting, approximating them to a 15-minute city concept.

2.4.1 Travel matrices

We iterate through all cities and perform a walking travel-time calculation, via the street network, from each hexagon's i centroid to every shop j in the dataset.

We use r^5py's wrapper for the R5 routing analysis engine, using as inputs the OSM network described in Section 2.3.2, the hexagons i as origins, the shops j as destinations, walking speed as the tool's default ($3.6kmh^{-1}$), maximum walking time of 960 minutes (16h), and only walking as travelling mode. The choice for this wrapper is due to an easy adaptation of the code in case future extensions include other travel modes, or a combination of them, which the engine allows. From the destination's side, we also extract the shop's category ("shop" = *) from OSM data, which accounts for over 340 classifications (not necessarily all present in the dataset, with possible unique values with combinations of them). In order to illustrate the outcome of these calculations, we select a central hexagon in Luxembourg City, highlighting the shops locations and the total travel time calculated from the hexagon to each one of them (Figure 2.4.1). The result of these calculations is a travel matrix per FUA with origin i IDs, destination j IDs, travel times between pairs of IDs (defined as t_{ij}), and shop category.

Figure 2.4.1: Travel times' calculation results from an illustrative hexagon in central Luxembourg towards nearby shops.



2.4.2 Accessibility metrics

The next step in the processing is to transform these travel times into meaningful accessibility indicators. Inspired by Tobler's First Law of Geography, the influence of a shop j on a hexagon i's accessibility should be inversely proportional to the distance between them. Transforming the travel times must, then, involve some function f with a negative slope ($f'(t_{ij}) \leq 0 \quad \forall \quad t_{ij}$). The strength at which this influence decays with distance is governed by what was conventionally called *friction*. The specific form of distance-decay functions must be carefully thought as it may result in misleading interpretations of impedance factors (Giannotti et al., 2022). We include, then, a friction parameter α into possible functions to transform the travel times. We analyse a series of transformations in line with literature for finally choosing a meaningful one for this analysis. The understanding of each one of them is done by highlighting which value a a shop j would add to a hexagon i's accessibility. The value a should necessarily be between 1 (whenever $t_{ij} = 0$) and 0 (whenever t_{ij} is so large it has negligible influence on a hexagon's accessibility).

We start by defining a negative exponential function $a^{exp}_{j|i}$ (Equation 2.1), which is commonly used in the literature but offers limited interpretability in this context. Next, we consider a power function, which faces issues when $t_{ij} \to 0$ as a tends towards infinity. To address this, we propose adding 1 to the travel time before transforming it ($a^{pow}_{j|i}$ - Equation 2.2), or alternatively, using a piece-wise function that caps a at 1 when $t_{ij} \leq 1$ ($a^{pw}_{j|i}$ - Equation 2.3). These adjustments help maintain the interpretability of the measure, reflecting the number of shops accessible per minute. This is essential as this research is embedded in the discussion of time-bounded accessibility towards shops (e.g., the 15-minute city).

Maintaining the metric in the sphere of shops per minute allows us to set time-specific bounds and threshold values b for t_{ij} above which a is taken as zero. This facilitates the exploration of scale-specific variations, for instance, identifying areas with more accessible shops within a 15-minute threshold. Meanwhile, it keeps the metric's interpretability consistent, especially under a constant friction parameter α. Although varying α could help achieve similar effects (e.g., higher friction resulting in more localised influences), keeping α constant enables more straightforward comparisons across the proposed metrics. Instead, we opt for a flexible choice of time-specific bounds b, keeping the simplicity of the interpretation. This leads us to our two remaining metrics: $a^{pow}_{j|i,b}$ (Equation 2.4), a case similar to $a^{pow}_{j|i}$ while excluding any shop above the threshold; and $a^{count}_{j|i,b}$ (Equation 2.5), a binary accessibility which indicates if shop j is under a b travel-time to hexagon i. The latter, when aggregated, provides a simple count of how many shops are accessible within b minutes from a hexagon's centroid, commonly called cumulative accessibility. The choices of which values of b are chosen in the analyses are further described in Section 2.4.3.

$$a_{j|i}^{exp} = e^{-\alpha t_{ij}} \tag{2.1}$$

$$a_{j|i}^{pow} = (1 + t_{ij})^{-\alpha} \tag{2.2}$$

$$a_{j|i}^{pw} = \begin{cases} 1 & \text{if } t_{ij} \leq 1 \\ t_{ij}^{-\alpha} & \text{otherwise} \end{cases} \tag{2.3}$$

$$a_{j|i,b}^{pow} = \begin{cases} (1 + t_{ij})^{-\alpha} & \text{if } t_{ij} \leq b \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

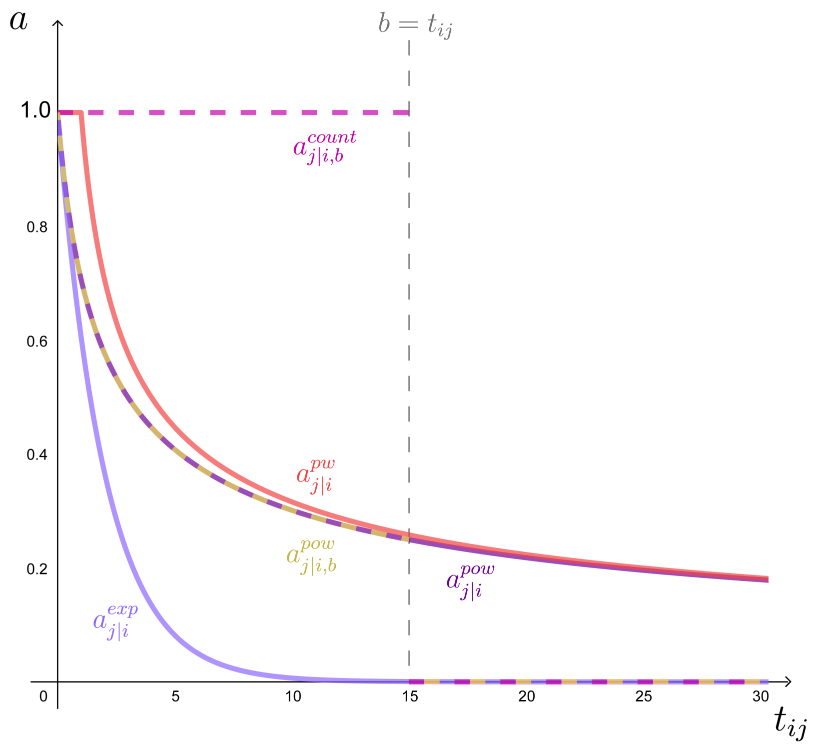
$$a_{j|i,b}^{count} = \begin{cases} 1 & \text{if } t_{ij} \leq b \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

We illustrate the differences in these different metrics by plotting the values achieved for a for $\alpha = 0.5$ in a schematic illustration (Figure 2.4.2). A higher (lower) value of α would indicate a stronger (weaker) friction, meaning all a values would decay quicker (less quick) with distance. We highlight that, apart from the $a_{j|i,b}^{count}$, the others function in a highly similar fashion, with their second derivatives being positive for all values of t (influence always decaying, but less strongly with distance). As argued before, we choose $a_{j|i,b}^{pow}$ to proceed with the analysis, allowing us flexibility to explore the multiple scales of the phenomenon, while keeping it easily interpretable.

In an applied setting, we illustrate these different accessibility metrics also by applying it to the context represented in Figure 2.4.3. Using the example of Luxembourg City, we select one hexagon in the central area of the city, and perform the calculations for each shop j with respect to that hexagon's centroid. We can see from Figure 2.4.3 that the nature of the metrics at shop level varies very little, except for sub-figure c, depicting a binary metric.

Given the chosen metric of $a_{j|i,b}^{pow}$, we define parameter $alpha = 1$, to keep the interpretability of the metric as "per unit time". We then aggregate the travel matrices per hexagon, summing over transformed travel times, depending on the bound chosen. These result in an aggregated accessibility indicator A per hexagon

Figure 2.4.2: Values achieved for a for any shop j as a function of travel time t for the different metrics proposed ($\alpha = 0.5$).



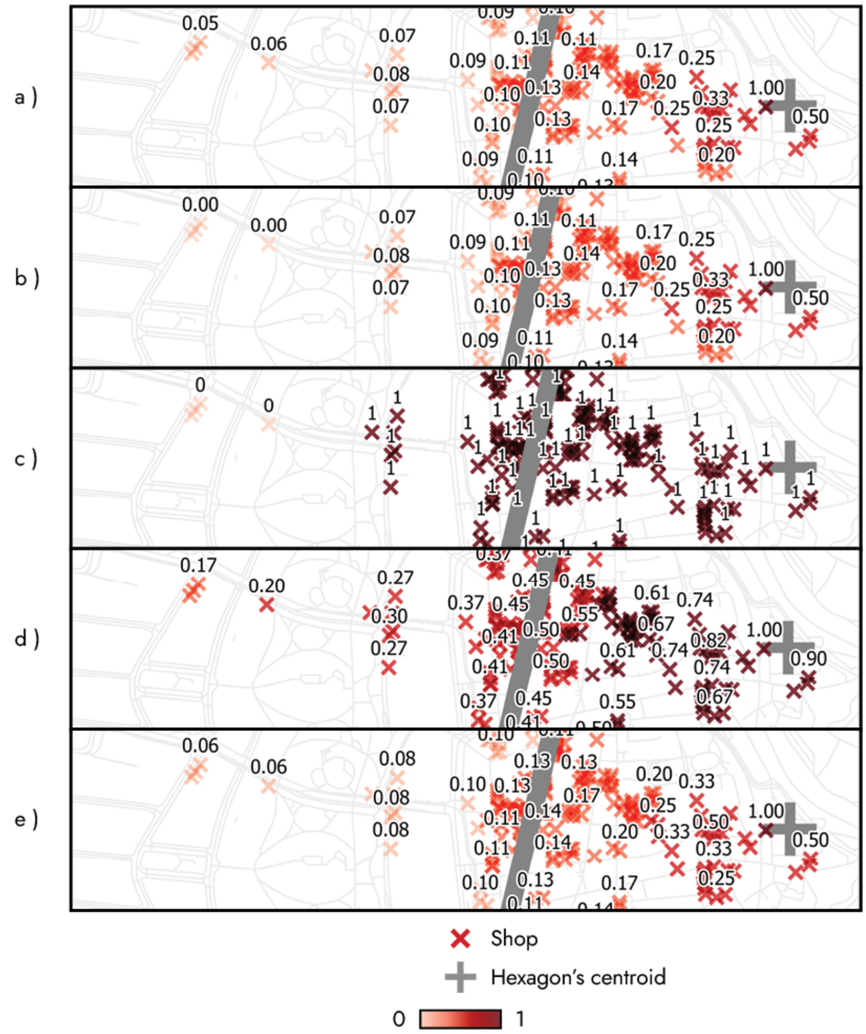
i, for a specific bound b, defined here as A_i^b (Equation 2.6), and are used in the subsequent steps of the analysis.

$$A_i^b = \sum_j^{J_c} (1 + t_{ij})^{-\alpha} \qquad \text{if } t_{ij} \leq b \tag{2.6}$$

2.4.3 Time-specific bounds

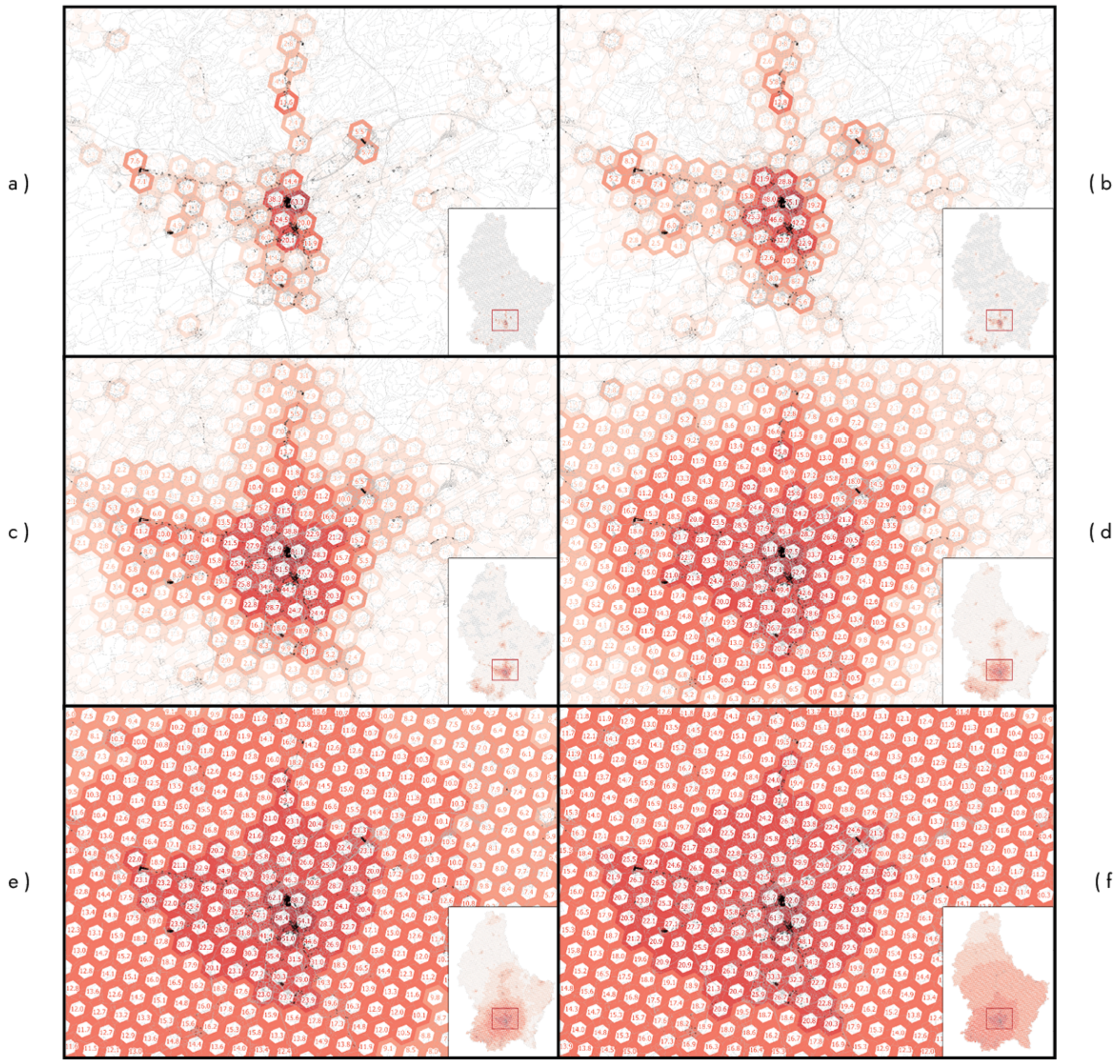
Before aggregating these travel times into the chosen accessibility indicator, we define a series of time-bounded thresholds to help make the analysis more granular. Once we aggregate travel times of maximum, for instance, 15 minutes, it highlights the more localised centralities of retail location. In this threshold, variations of accessibility between hexagons are more pronounced, and the resulting values, as a scalar surface, seem more rough. On the other hand, when we do not limit the travel time with thresholds, we achieve a more "smoothed out" surface, i.e., more homogeneous. These bounds, as mentioned before, serve as a proxy for how locally central each hexagon is. The higher the bound, the harder it is for a hexagon to

Figure 2.4.3: Resulting accessibility metric a for each shop j w.r.t. the reference hexagon. Sub-figure a) $a^{pow}_{j|i}$, for $\alpha = 1$. Sub-figure b) $a^{pow}_{j|i,b}$, for $\alpha = 1$, $b = 15$. Sub-figure c) $a^{count}_{j|i,b}$ for $b = 15$. Sub-figure d) $a^{exp}_{j|i}$ for $\alpha = 0.1$. Sub-figure e) $a^{pw}_{j|i}$ for $\alpha = 1$.



remain accessible (as compared to its neighbours), as it captures the influences of shops farther away. Departing from the 15-minute city discussion, we choose the bounds as multiples of 15, doubling at each step: 15 minutes, 30 minutes, 60 minutes, 120 minutes and 240 minutes. Considering a walking travel time of 240 minutes seems like an exaggeration at first, but it can be seen as a proxy for centrality arising from other, faster, means of transportation (e.g., cycling, transit, driving). Despite the network not being exactly the same, at that level of aggregation, it is a close approximation. We add a sixth, unbounded indicator ($b \to \infty$), representing the overall accessibility per hexagon (A_i^{∞}). We iterate these calculations for the travel matrices of every city and we illustrate the method, again, with the example of Luxembourg City in Figure 2.4.4. We can see the rougher scalar surface for lower bounds in sub-Figure 2.4.4-a, while the more smoothed-out one in sub-Figure 2.4.4-f.

Figure 2.4.4: Resulting values for accessibility indicators, exemplified in Luxembourg, per hexagon. Sub-figure a) A_i^{15}. Sub-figure b) A_i^{30}. Sub-figure c) A_i^{60}. Sub-figure d) A_i^{120}. Sub-figure e) A_i^{240}. Sub-figure f) A_i^{∞}.



2.4.4 Aggregated accessibilities

The next step in the analysis is to aggregate the calculated hexagon-level accessibility indicators per city. We propose this aggregation in two steps: first, we define a simple average of the accessibility indicator per city (Equation 2.7); secondly, we weigh the indicator by the population of each hexagon. Since each hexagon has roughly the same area, weighting by population has the same effect as weighting by population density. In this analysis, a population-weighted accessibility represents the number of shops per minute the average citizen reaches. This allows us to draw comparisons between different cities for each threshold. The population-weighted accessibility per city c, per threshold b, is defined in Equation 2.8 as $\bar{A}_c^b$. To visualise these results, we choose first a distribution plot for the unweighted average; and for the population-weighted average a series of maps, one per threshold, as well as a parallel coordinates plot to compare cities as we increase the threshold.

$$A_c^{\infty} = \frac{1}{I_c} \sum_i^{I_c} A_i^b \tag{2.7}$$

$$\bar{A}_c^b = \frac{1}{N_c} \sum_i^{I_c} n_i A_i^b \tag{2.8}$$

2.4.5 15-Minute Tightness

To evaluate the different cities in terms of how they adhere to a 15-minute city concept we choose two strategies. Firstly, we rank the cities in terms of their 15-minute population-weighted accessibilites. This already discloses how well connected the average citizen is to shops within 15 minutes from their place of residence. We aggregate these values by country, weighting by city population N_c. Secondly, we plot the 15-minute population-weighted accessibility ($\bar{A}_c^{15}$) against the 240-minute population weighted accessibility ($\bar{A}_c^{240}$), we fit a linear function to the data, then calculate the residuals - how far from the fit each city is. This is expected to indicate a level of tightness of a city's retailscape, as it compares a walkable 15-minute accessibility to a non-walkable accessibility, as a 4h walking time is equivalent to approximately a 30-minute drive. Cities above the trend line are expected to be less tight, and the ones below to have a tighter retailscape (closer to the 15-minute city conceptual model).

2.5 Results

In this section, we explore different dimensions of accessibility to retail to better understand regional variation between population and commercial density. We use a combination of descriptive statistics, population-weighted accessibility metrics, and finally, evaluations of 15-minute tightness.

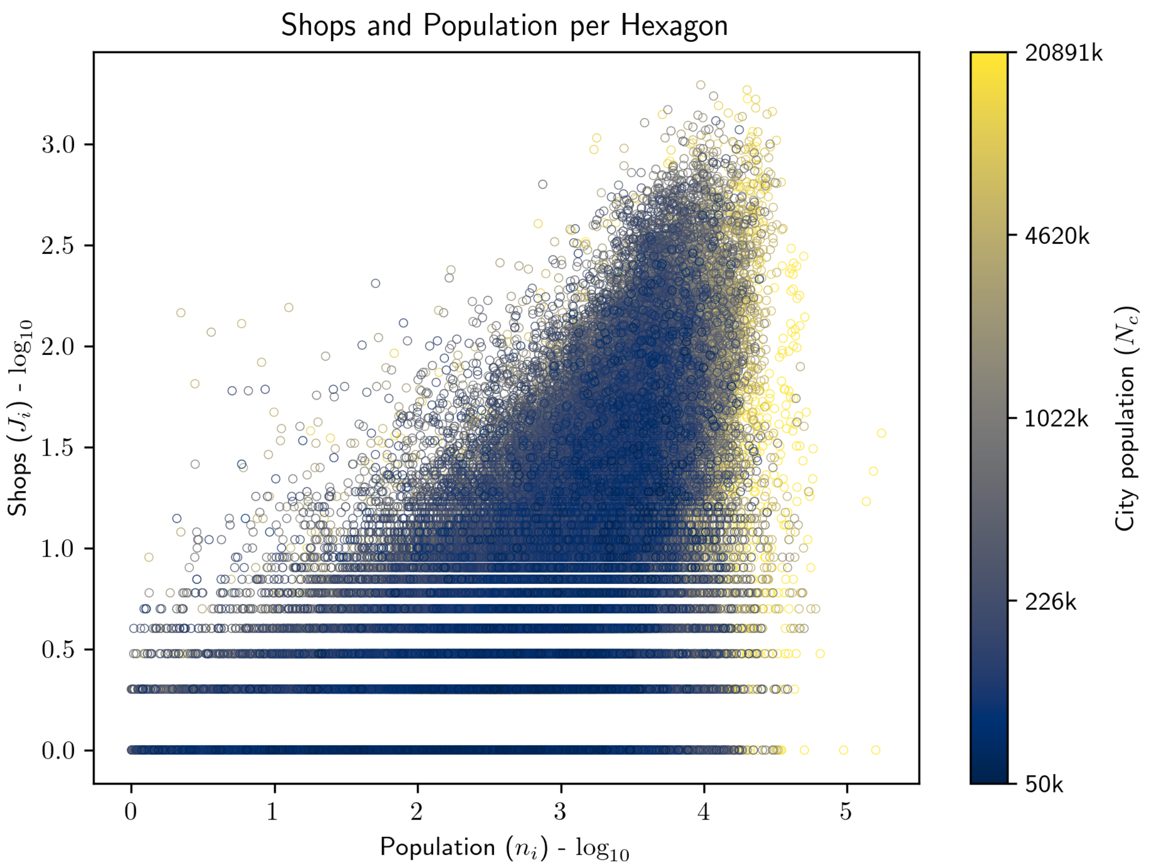
2.5.1 Descriptives

As a first step, to answer the first proposed research question, we descriptively plot the extracted number of shops and population per hexagon against each other (Figure 2.5.1), colour-coding them by the total city population. We find that a low number of shops can be found in a wide range of population values at hexagon-level, but a high number of shops is only found within higher population ranges. This may seem trivial at first, but it is not: recent trends of segregated land-use and shopping-centre dissemination could have detached shop clustering patterns from population centres, but these findings suggest otherwise. It indicates that population concentration still dominates over shop locations. By colour-coding the hexagon by its respective city-level population, we notice that this pattern is not restricted to bigger cities, but we detect that high population level in a hexagon does correlate with city size (Easting in Figure 2.5.1). In other words, hexagons with the highest population values (densest) are most likely found within bigger cities, confirming an initial hypothesis that larger urban areas do experience also higher population densities, to some extent. Since this is a possible reason for the superlinearity between population size and total number of shops, as detected in Chapter 1, it is worht highlighting.

2.5.2 Aggregated accessibilities

Our analysis advances to aggregating accessibility to retail at a city level, to answer the second research question. We first explore the unweighted, unbounded accessibility indicator, averaged per city. We highlight, by plotting the distribution of the accessibility per hexagon per city (Figure 2.5.2), the number of zero or near-zero values distorting the indicator, indicating a strong bias regarding the format and definition of the Functional Urban Area itself. By weighing per population,

Figure 2.5.1: Scatterplot between log of number of shops (J_i) and population (n_i) at hexagon level.

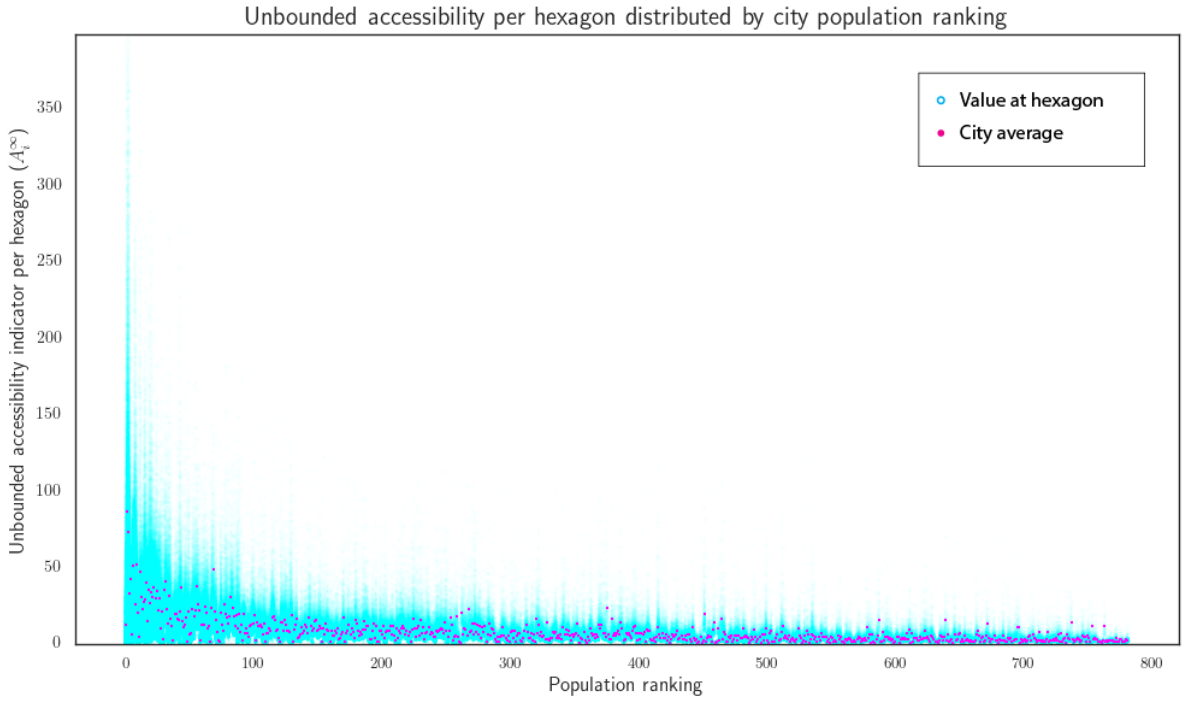


we attempt to address this bias, making the indicator readable and interpretable regardless of the different definitions of urban areas' boundaries. In other words, low (high) values of accessibilities in low- (high-) density areas affect less (more) people and may negatively (positively) contaminate the simple arithmetic average when FUA boundaries include more (less) empty spaces.

Unweighted accessibilities

To illustrate the variation in retail accessibility across hexagons within a city, and across different cities, we plot the range of unbounded, unweighted accessibility indicators ranked by city population (Figure 2.5.2). We notice that the city averages, in this case, are more consistent with their size, exhibiting a declining trend as we move towards the right on the chart (towards smaller cities). We highlight that cities with expansive urban areas tend to show a lower average of unweighted accessibility, due to the inclusion of low-density outskirts. This further justifies the importance of considering population distribution when averaging accessibility indices, as vast areas with low population densities may negatively skew the results. The unweighted accessibility, thus, fails to capture the actually perceived accessibility to shops by local residents, which is addressed in the weighted accessibilities, described in the next section.

Figure 2.5.2: Unbounded accessibility per hexagon (A_i^{∞}), ranked by city size.



Population-weighted accessibilities

Our analysis then advances to assessing accessibility to retail weighted by population. This approach quantifies how accessible shops are to the average city dweller, using travel-time thresholds of 15, 30, 60, 120, and 240 minutes. This way, we offer a quantified assessment of an important aspect of urban convenience. We present the summary of results in Figure 2.5.3. We highlight that, for very local accessibility (15- and 30-minute thresholds), smaller cities in the countryside of Mediterranean countries, such as Spain and Greece, are highlighted as having very strong accessibilities. As we move towards higher thresholds, central more populous cities take the lead, such as Paris, London and Berlin. Smaller cities of Türkiye and Eastern European countries, such as Romania and Bulgaria, still figure among the lowest accessibilities even as we move towards higher thresholds. It is important to highlight that this metric is still sensitive to the area of the FUAs. Cities with larger areas scale more intensively as we increase the threshold. We also plot the top and bottom 5 cities for each threshold in a parallel coordinates plot, illustrated in Figure 2.5.4.

Figure 2.5.3: Resulting values for population-weighted accessibilities in Europe, per city. Sub-figure a) $\bar{A}_c^{15}$. Sub-figure b) $\bar{A}_c^{30}$. Sub-figure c) $\bar{A}_c^{60}$. Sub-figure d) $\bar{A}_c^{120}$. Sub-figure e) $\bar{A}_c^{240}$. Sub-figure f) $\bar{A}_c^{\infty}$.

a) (b

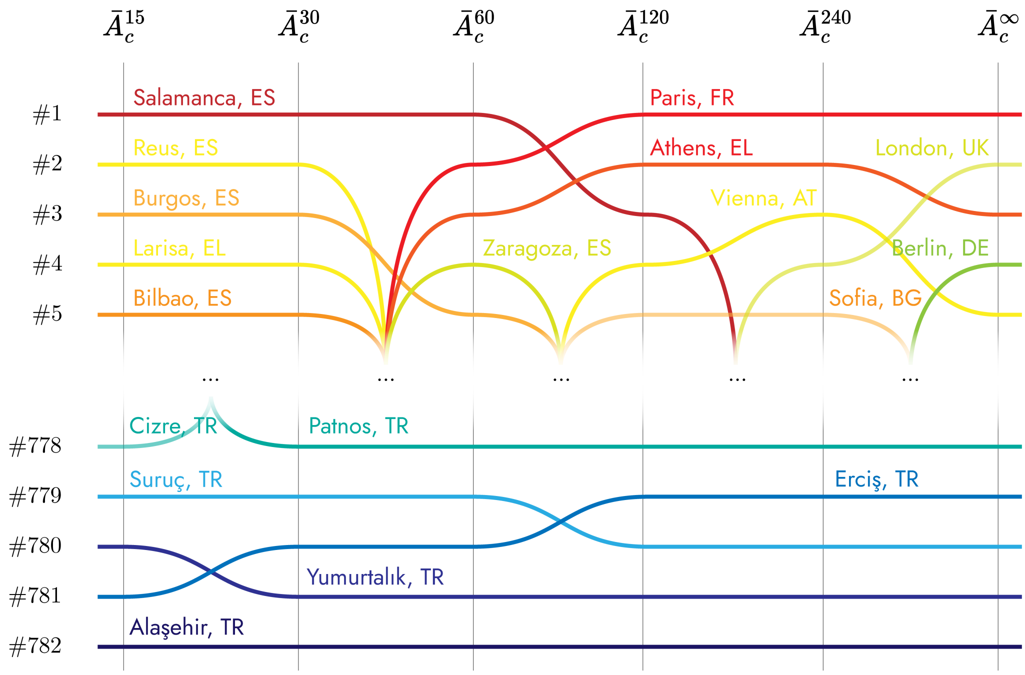
c) (d

e) (f

2.5.3 Country-level aggregations

We then aggregate the 15-minute accessibilities by country, again weighting by population for each FUA. The selection of this specific time-bound gives us an initial overview of the adherence to a 15-minute city concept by European cities already, which is further refined in Section 2.5.4. Figure 2.5.5 shows that, aligned to what was detected in an initial visual inspection with Figure 2.5.3, Greece and Spain figure among the highest 15-minute accessibilities across Europe. Following are Austria, Albania, France and Bulgaria, respectively. France and Austria's numbers,

Figure 2.5.4: Parallel coordinates plot with the rank of top 5 and bottom 5 cities for each accessibility indicator threshold.

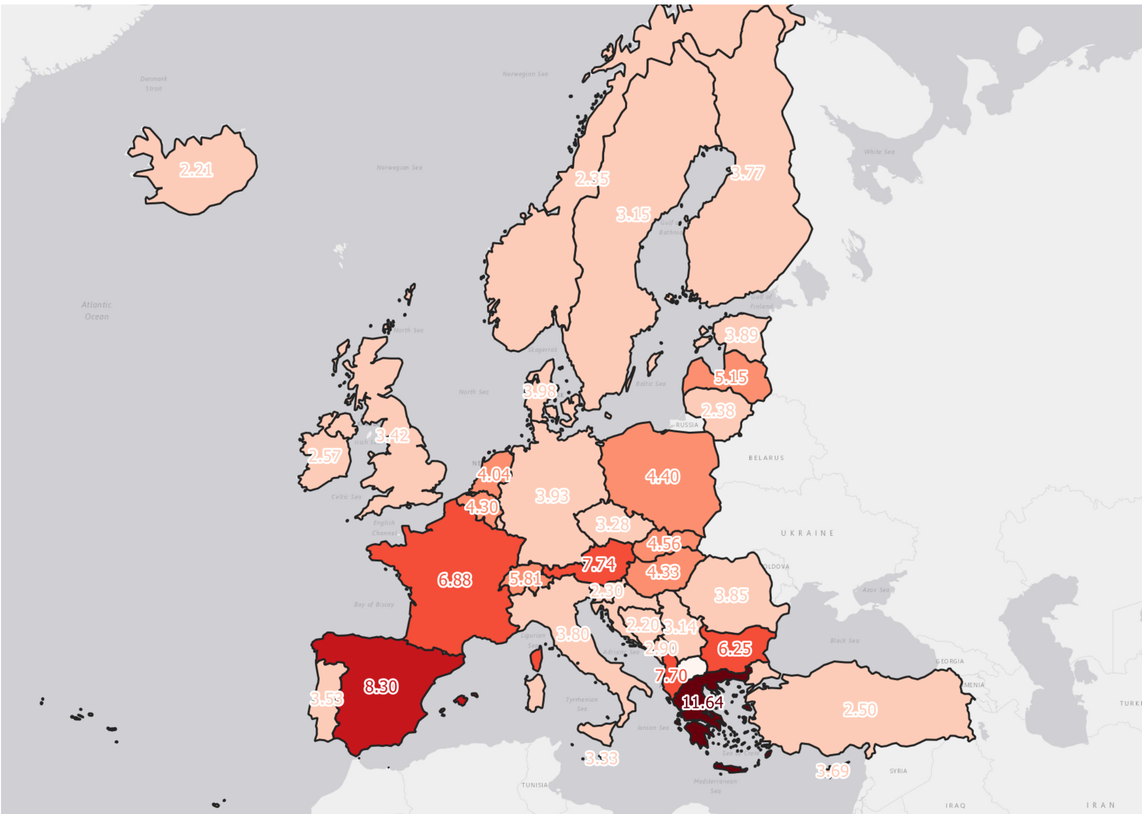


in particular, are detected to be pulled by their main city's accessibility: both Paris and Vienna have a much higher value than their country's average without them. This could be explained partly by historical reasons, as both Paris and Vienna are historically cultural strongholds of former large empires.

2.5.4 15-Minute Tightness

Finally, we evaluate what we call the 15-minute tightness (15'T) by comparing accessibility to retail within two time-bounded thresholds: 15 minutes and 240 minutes. As expected, the two accessibility indicators are strongly correlated, as the larger threshold necessarily includes the smaller one, by design. By plotting them against each other, we fit a linear function as a trend-line and calculate the vertical distance in the plot from all values to this trend-line. Cities located below this line indicate a closer-to-linear relation between the two accessibilities. In other words, in these cities, the reached accessibility does not change that much as we increase the threshold. This means the retail accessed within a 15-minute walk does not differ much from that accessed within 4 hours (or a 30-minute drive). We choose this metric to indicate the level of tightness of cities. On the other hand, cities above this line indicate the opposite phenomenon, where accessibility increases a lot when

Figure 2.5.5: Aggregated 15-minute accessibilities ($\bar{A}_c^{15}$) per country. Aggregations are weighted by each FUA's population.



you start considering higher thresholds. These are defined as the least tight. This is shown in Figure 2.5.6.

With this definition in mind, we then analyse a map of these values, highlighting where are the most and least tight cities detected. Figure 2.5.7 maps these values. We can see that Spanish cities, already highlighted as highly accessible within the 15-minute threshold, are also highly tight. This means that their accessibility at lower travel-time thresholds is high even when compared to a higher threshold. Other areas visually detected with higher concentrations of tight cities are southern Italy, and coastal cities in Albania and Greece. These Mediterranean cities perhaps offer a historical city-centre layout that has allowed them to have a tighter offer of retail, not extending much beyond them.

2.6 Conclusion

This work analysed retail accessibility across European cities, highlighting regional variations and connecting the concepts of retail distribution, population density, and retail tightness. The first contribution of this research is its scale, including

Figure 2.5.6: Definition of tightness with 15-minute accessibility ($\bar{A}_c^{15}$) plotted against 240-minute accessibility ($\bar{A}_c^{240}$).

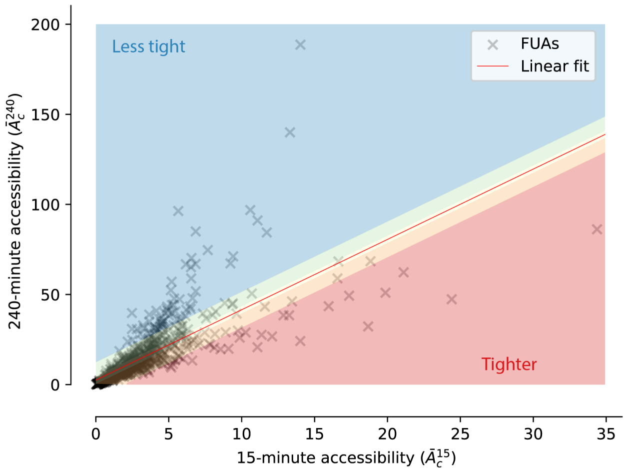
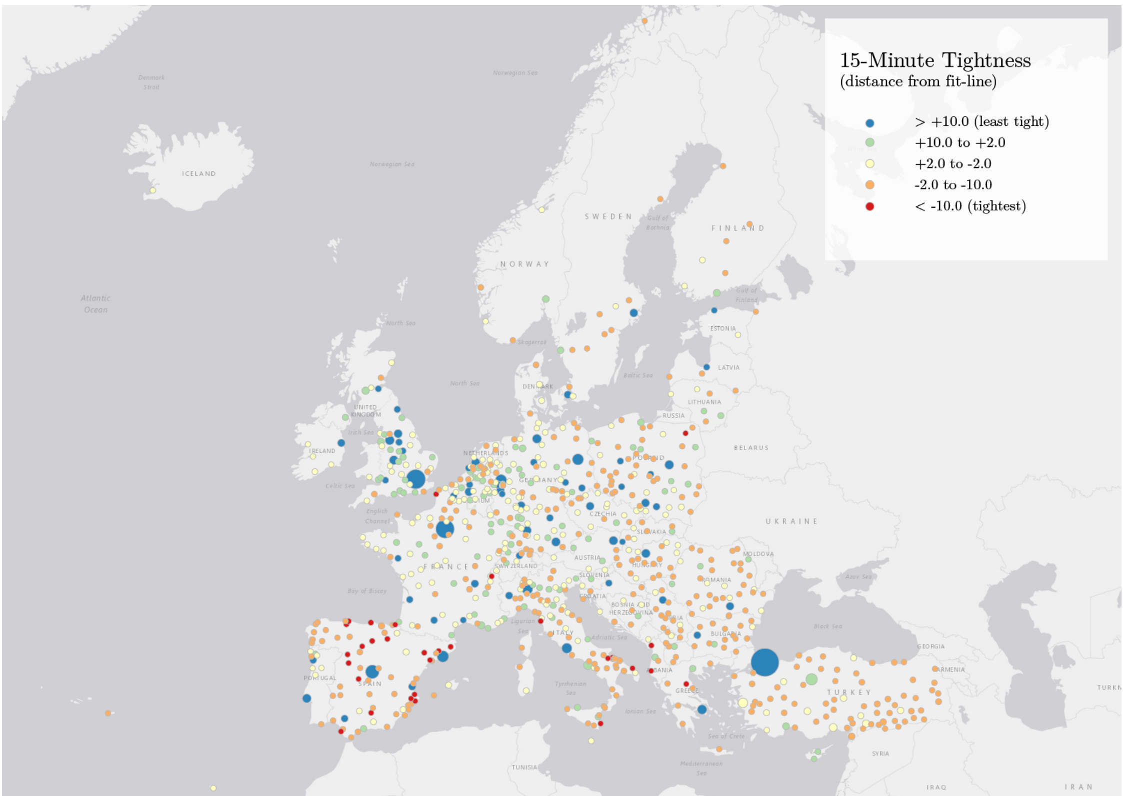


Figure 2.5.7: Map of 15-Minute Tightness (15'T) of the different European cities.



nearly 800 cities across Europe. By systematically applying cumulative accessibility methodologies to each FUA, this study allows for cross-city comparisons and country-level aggregation. By doing so, it expands the scope of cumulative accessibility research from localised studies to a broad continental analysis. Moreover, this study addresses in-depth a specific dimension in the 15-minute-city discussion: the access to shops. While the accessibility literature often includes access to opportunities (e.g., jobs), services, and amenities, this study's contribution lies in its focus specifically on retail access. This allows for a better understanding of how retail location patterns relate to population distribution, both mediated by the infrastructure networks.

We also find that the patterns of retail concentration found in Europe are still very aligned with how people concentrate in space. We detect a pattern of nestedness, indicating that a higher density of shops is only found in areas of higher population density, but not the other way around. This suggests that, despite recent trends in suburbanisation of shopping opportunities (with shopping centres, and isolated suburban shopping areas), the dominating trend is still supply following demand. A higher concentration of people does not necessarily mean a higher density of shops, but the latter practically only happens with the former. This also aligns with the findings that indicate that shopping concentration patterns scale superlinearly with population, suggesting the effect of agglomeration economies in place.

The study's findings indicate that smaller cities in the Mediterranean countries, notably Spain and Greece, exhibit exceptionally high accessibilities within short travel-time thresholds. This phenomenon could be attributed to the dense, historically compact urban cores commonly found in these areas. Conversely, as we extend the travel-time threshold, larger metropolitan areas such as Paris, London, and Berlin appear with the highest accessibility, suggesting already a later observed superlinear correlation between city size and access to shops.

It is important to highlight, at this point, a key limitation of this research. The fact that we rely on data from OSM, which is voluntarily sourced by active users, might limit the extension of this sort of conclusion. It could be that smaller cities have, instead, a disproportionately lower number of users contributing to the data available in OSM. Further robustness should be needed to confirm it, but it could be that we are, in fact, capturing an agglomeration effect towards the number of OSM active users, instead of the number of shops per se. This would mean that larger cities, with disproportionately higher number of active contributing users, have a more accurate representation of shops, while smaller cities are underrepresented.

Finally, the study of tightness through accessibility metrics at varying thresholds attempts to qualify urban compactness, especially in the sphere of retail, a concept commonly found in literature as a desirable feature cities should have when planning for sustainability. Cities identified as tight offer consistent retail access regardless of the travel-time threshold, indicating a potentially more sustainable urban configuration, where citizens do not need to rely on the private automobile regardless of their destination. This attribute is predominantly observed in Mediterranean cities, possibly reflecting their traditional urban layouts that naturally encourage retail proximity.

References

Araldi, A., & Fusco, G. (2019). Retail Fabric Assessment: Describing retail patterns within urban space. *Cities*, *85*, 51–62.

Barbieri, L., D'Autilia, R., Marrone, P., & Montella, I. (2023). Graph Representation of the 15-Minute City: A Comparison between Rome, London, and Paris. *Sustainability*, *15*(4), 3772.

Bartzokas-Tsiompras, A., & Bakogiannis, E. (2023). Quantifying and visualizing the 15-Minute walkable city concept across Europe: A multicriteria approach. *Journal of Maps*, *19*(1), 2141143.

Bartzokas-Tsiompras, A., Bakogiannis, E., & Nikitas, A. (2023). Global microscale walkability ratings and rankings: A novel composite indicator for 59 European city centres. *Journal of Transport Geography*, *111*, 103645.

Becerra Valbuena, L. G. (2013). Aproximaciones microeconómicas en la Teoría de los Lugares Centrales de Christaller. *Ensayos sobre Política Económica*, *31*(70), 67–120.

Boisjoly, G., & El-Geneidy, A. M. (2017). The insider: A planners' perspective on accessibility. *Journal of Transport Geography*, *64*, 33–43.

Cachinho, H. (2014). Consumerscapes and the resilience assessment of urban retail systems. *Cities*, *36*, 131–144.

Christaller, W. (1966). *Central Places in Southern Germany*. Prentice-Hall.

d'Aspremont, C., Gabszewicz, J. J., & Thisse, J.-F. (1979). On Hotelling's "Stability in Competition". *Econometrica*, *47*(5), 1145.

Dahn. (2021). *H3-pandas: H3 indexing for pandas dataframes* [Available at: https://github.com/DigitalGlobe/h3-pandas].

Efeoglu, H., Joutsiniemi, A., & Mozuriunaite, S. (2023). Exploring the plot patterns of the retail landscape: The case of the Helsinki Metropolitan area. *Environment and Planning B: Urban Analytics and City Science*.

Ferrer-Ortiz, C., Marquet, O., Mojica, L., & Vich, G. (2022). Barcelona under the 15-Minute City Lens: Mapping the Accessibility and Proximity Potential Based on Pedestrian Travel Times. *Smart Cities*, *5*(1), 146–161.

Fujita, M., Krugman, P., & Mori, T. (1999). On the evolution of hierarchical urban systems. *European Economic Review*, *43*(2), 209–251.

Giannotti, M., Tomasiello, D. B., & Bittencourt, T. A. (2022). The bias in estimating accessibility inequalities using gravity-based metrics. *Journal of Transport Geography*, *101*, 103337.

Graells-Garrido, E., Serra-Burriel, F., Rowe, F., Cucchietti, F. M., & Reyes, P. (2021). A city of cities: Measuring how 15-minutes urban accessibility shapes human mobility in Barcelona. *PLOS ONE*, *16*(5), e0250080.

Hidalgo, C. A., Castañer, E., & Sevtsuk, A. (2020). The amenity mix of urban neighborhoods. *Habitat International*.

Hogba, E., & Yiran, G. (2024). Analysing the spatial pattern of retail store locations in the urban environment of accra using GIS. *GeoJournal*, *89*(3).

Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, *39*(153), 41–57.

Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020, July). *Geopandas/geopandas: V0.8.1* (Version v0.8.1). Zenodo.

Lemoy, R., & Caruso, G. (2020). Evidence for the homothetic scaling of urban forms. *Environment and Planning B: Urban Analytics and City Science*, *47*(5), 870–888.

Moreno, C., Allam, Z., Chabaud, D., Gall, C., & Pratlong, F. (2021). Introducing the “15-Minute City”: Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities*, *4*(1), 93–111.

Olivari, B., Cipriano, P., Napolitano, M., & Giovannini, L. (2023). Are Italian cities already 15-minute? Presenting the Next Proximity Index: A novel and scalable way to measure it, based on open data. *Journal of Urban Mobility*, *4*, 100057.

Openshaw, S., & Veneris, Y. (2003). Numerical Experiments with Central Place Theory and Spatial Interaction Modelling. *Environment and Planning A: Economy and Space*, *35*(8), 1389–1403.

Pereira, R., Banister, D., Schwanen, T., & Wessel, N. (2019). Distributional effects of transport policies on inequalities in access to opportunities in Rio de Janeiro. *Journal of Transport and Land Use*, *12*(1), 741–764.

Ribeiro, F. L., & Rybski, D. (2023). Mathematical models to explain the origin of urban scaling laws. *Physics Reports*, *1012*, 1–39.

Schiavina, M., Freire, S., & MacManus, K. (2023, April). GHS-POP R2023A - GHS population grid multitemporal (1975-2030).

Taylor, P. J., & Hoyler, M. (2021). Lost in plain sight: Revealing central flow process in Christaller’s original central place systems. *Regional Studies*, *55*(2), 345–353.

Tomasiello, D. B., Herszenhut, D., Oliveira, J. L. A., Braga, C. K. V., & Pereira, R. H. M. (2023). A time interval metric for cumulative opportunity accessibility. *Applied Geography*, *157*, 103007.

Ulloa-Leon, F., Correa-Parra, J., Vergara-Perucich, F., Cancino-Contreras, F., & Aguirre-Nuñez, C. (2023). “15-Minute City” and Elderly People: Thinking about Healthy Cities. *Smart Cities*, *6*(2), 1043–1058.

Vale, D., & Lopes, A. S. (2023). Accessibility inequality across Europe: A comparison of 15-minute pedestrian accessibility in cities with 100,000 or more inhabitants. *npj Urban Sustainability*, *3*(1), 1–13.

Wachs, M., & Kumagai, T. G. (1973). Physical accessibility as a social indicator. *Socio-Economic Planning Sciences*, *7*(5), 437–456.

Chapter 3

Bringing economic complexity to the intra-urban scale: the role of services in the urban economy of Belo Horizonte, Brazil

This study explores the formation of economic complexity within a city from the Global South, during 2011-2019. It proposes an expanded interpretation of the Economic Complexity Index (ECI) to be applied at the intra-urban context of Belo Horizonte, Brazil, focusing on three different spatial levels of analysis (i.e., local, neighbourhood, and community levels). By applying the index to these three levels, instead of regional or national administrative boundaries commonly used in literature, this study contributes to approximating the observation of economic complexity to the actual geographical scales at which economic interactions take place, allowing for intra-urban comparisons. The proposed ECI includes the service economy, amenities, and retail, in addition to commonly observed manufacturing industry. Methodologically, this case study introduces the Urban Economy Space network diagram to the expanded ECI as an effort to holistically consider all economic sectors happening in a city. The main findings are twofold. First, the city services classified as more complex by the ECI are aligned with the theory of post-industrial economic

activities: financial, telecommunications, scientific and technical services, etc. Second, government-led institutions such as healthcare facilities, higher education institutions, etc., appear on the top tier of economic complexity, indicating that local and national governments can contribute to complexifying local economies.

3.1 Introduction

The process of innovation happens essentially within cities, when new labour is added on top of old labour and new business categories arise as breakaways from existing ones (Jacobs, 1970). This concept of innovation leads to the idea of economic development as a process being fairly linked to the diversification of a city's economic activities. Economic diversification also plays a role in enhancing a region's resilience, i.e., its capacity to withstand external shocks (Frenken et al., 2007; Sprague, 2018; Xiao & Drucker, 2013). Despite urban areas being crucial for economic development, current literature in geographies of innovation tends to focus on firms and entrepreneurial activities, at the regional or national scales (Florida et al., 2017), thus not usually addressing urban or intra-urban scales (Adler et al., 2019).

Recent economic geography literature points to economic complexity as a key ingredient for analysing an area's composition of economic activities (Bishop & Mateos-Garcia, 2019; Burlina & Antonietti, 2020). Economic complexity of countries is also positively associated with their levels of economic development (Hidalgo, 2015). As a measure at a national scale, the Economic Complexity Index (ECI) (Hidalgo & Hausmann, 2009) classifies products exported by countries, ranking their economies from more to less complex. The ECI has also been applied to regional economies (Bishop & Mateos-Garcia, 2019; Burlina & Antonietti, 2020), but is seldom used for understanding and comparing the economies of different areas within a city, nor applied to urban services, amenities, and retail, in extension to manufacturing industries or in place of exported products. This paper attempts to address this gap, applying the ECI to intra-urban areas, using as case study a city of the Global South.

3.1.1 Complexity and the city's economy

Economic diversification as a process is commonly seen as opposed to economic specialisation (Batty, 2017). While the former is considered a result of innovation led by knowledge spillovers in cities, the latter points out that agglomeration economies lead to ever-more specialised industries in a highly entrepreneurial environment (Burlina & Antonietti, 2020; Faggio et al., 2020). The classical work of Jacobs (1970) credits the rise in economic diversity of cities to recurrent episodes of "import replacement". New industries may create new comparative advantages for products or processes in a city, where they did not previously happen, generating diversification. This is one of the reasons cities may abandon previous production processes, generating specialisation. This can lead to the emergence of related services and products, defined in literature as related variety (Bond-Smith & McCann, 2020; Frenken et al., 2007), that can again increase economic diversity and, in turn, foster further specialisation. These possible recurring processes of diversification and specialisation show that these two concepts do not necessarily oppose each other, but a successful economy may combine them in a harmonic way, leading to higher economic competitiveness in cities (Hong & Xiao, 2016).

Feedback processes across different scales, similar to the one described, is a key characteristic of complex systems studies (Salvati et al., 2015). Applied to whole economies, the study of economic complexity emerges as a relatively recent field, trying to link various socioeconomic phenomena to these non-linear dynamics of complex systems (Balland et al., 2022). Higher economic complexity has been linked to reduced greenhouse gas emissions (Mealy & Teytelboym, 2020; Romero & Gramkow, 2021), reduced shares of the shadow economy (Ha et al., 2021; Nguyen, 2022), reduced income and regional inequalities (Marco et al., 2022; Zhu et al., 2020), and a stronger income convergence between countries (Gala et al., 2018). Bringing it closer in scale to the object of this research is possible since urban economies can also be characterised as complex systems, i.e., they are composed of heterogeneous agents in a variety of groups, acting in different times and spaces, incurring in non-linear patterns, generating unexpected outcomes (Burlina & Antonietti, 2020).

The idea of economic complexity in cities is also pervaded by the presence of knowledge-intensive, innovative products and services (Balland & Rigby, 2017). Large and well-connected cities tend to disproportionately concentrate innovation within its firms and production processes (Balland et al., 2020), usually associated with the post-industrial knowledge economy (Murdoch, 2018; Sassen, 2005). The presence of creative services in cities also contributes to higher employment growth and new business formation (Boschma & Fritsch, 2007). On the other hand, transition towards an intangible digital economy poses challenges for cities, such as rising structural unemployment with the gradual abandonment of mass-production systems (Bertani et al., 2021). Furthermore, Bettencourt et al. (2007) highlight the need for cities to continuously accelerate innovation cycles, essentially carried out within the sphere of knowledge-centred services and processes, in order to avoid stagnation or collapse. Addressing the digital transitions without leaving its citizens behind, while still avoiding the constant risk of stagnation and collapse, is a pressing dilemma for planners worldwide.

The knowledge-intensive service economy, essential for the rise and maintenance of this innovative environment, is embedded with consultancy firms, law firms, financial services, the creative sector, research and development (R&D). However, these as economic sectors are not often included in analyses of economic complexity. Other studies may include neighbourhood amenities (i.e., restaurants, hotels, bars, cafés) in analyses of economic relatedness and complexity (Hidalgo & Castañer, 2015; Hidalgo et al., 2020), but still leaving aside the previously mentioned service sectors, as well as retail. Therefore, there is a need for a holistic approach including multiple economic sectors happening within cities (manufacturing industries, the knowledge economy, amenities) to understand and study the entirety of a city's complex economic structure.

3.1.2 Quantifying economic complexity: the ECI

One way to quantify the complexity of a city's economic structure is by using the ECI (Hidalgo & Hausmann, 2009). The index originally assigns complexity values to both exported products and countries, representing the amount of crystallised knowledge involved in a product, and the presence of the right set of capabilities in a country for producing such products (Hidalgo, 2015). Since transmission of knowledge and capabilities has a strong geographic component, it has been described as being 'spatially sticky' (Balland & Rigby, 2017). Moreover, the ECI has also been applied to regional economies, by considering levels of employment in specific industry categories (Bishop & Mateos-Garcia, 2019; Burlina & Antonietti, 2020; Cicerone et al., 2019), producing aggregated results for regional administrative boundaries. These aggregations, however, often fail to capture internal variations of the regions analysed, also ignoring key differences in local economic structures within the same city. By applying it to intra-urban areas these questions are addressed in this paper.

The seminal work regarding the Economic Complexity Index is Hidalgo and Hausmann (2009). From then, a series of multiple approaches have been implemented, having the field evolved in different directions, either towards studies of innovation via patent data, towards industrial organisation and industrial composition in a Schumpeterian approach (Sciarra et al., 2020). Complexity measurements have been associated with economic performance (Gala et al., 2018), environmental performance (Romero & Gramkow, 2021), inequality (Marco et al., 2022). Refer to the work of Balland et al. (2022) for more detailed descriptions, references, and examples.

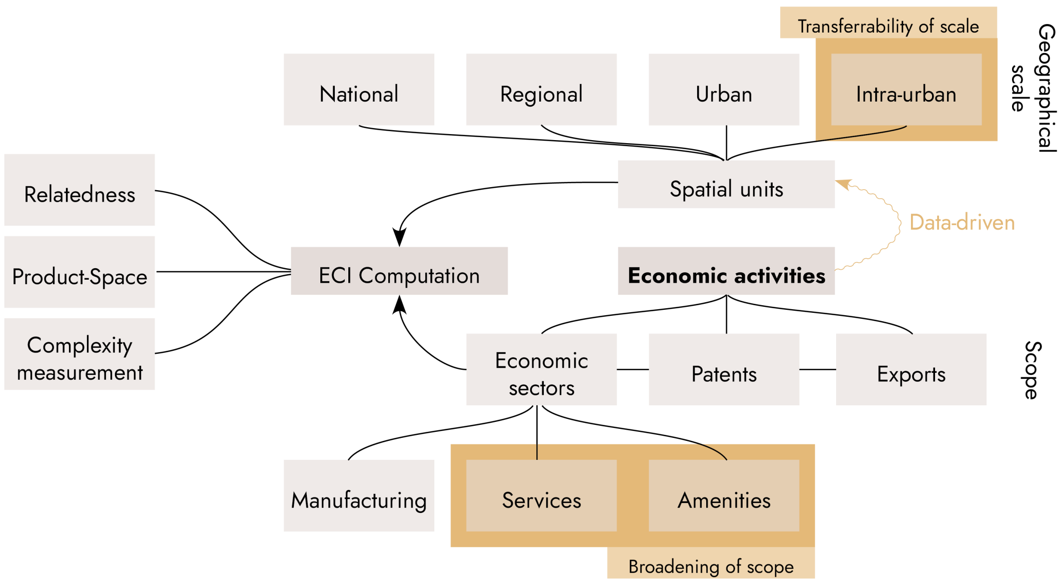
3.1.3 Research aims

The ECI is often applied to products exported by countries or regional labour structures, but seldom to urban or intra-urban economies. This research aims at understanding the structure of a city's economy, as measured by the ECI, and its derived relatedness matrices, and product-space networks, at different spatial scales. This allows to test the ECI's broadening of scope, by holistically considering all economic sectors in place of solely the commonly observed manufacturing industry; as well as the ECI's transferability of scale, by applying the index to intra-urban areas, using multi-scalar, data-driven spatial units.

3.1.4 Conceptual framework

The following conceptual framework (Figure 3.1.1) summarises the current state-of-the-art of the literature in the topic, as well as highlights the main theoretical contributions of this research, beyond the shift of focus towards an emerging economy city. The presence of economic activities in a given (or chosen) geographical scale is the base object of analysis in any study about economic complexity, and the reading of this framework should start from it. However, the specific scope and object of analyses vary from field to field, or from research to research. It is common to find analyses that focus on exports (Poncet & Starosta de Waldemar, 2013), patent data (**balland2019a**), or the presence of certain economic sectors, such as occupational data from STEM categories (Lo Turco & Maggioni, 2022; Mealy & Teytelboym, 2020). The presence of economic activities by sector is often seen from the perspective of the manufacturing sectors (Dosi et al., 2022). The combination of these scopes with the spatial units chosen allows for the computation of the Economic Complexity Index, as well as all its derived measurements, such as relatedness matrices, the Product-Space, and the complexity measurement itself. At last, but not least, we consider the data-driven definition of spatial units to be an extra contribution of this work to the literature.

Figure 3.1.1: Conceptual framework, highlighting the main contributions of this research to current literature in the topic.



3.2 Methods

This section starts by defining the study area and the importance of such a study being done in its context. It follows with the description of the datasets used, and the delimitation of spatial units using a data-driven, bottom-up approach. Finally, it explains the calculation of the ECI index for the spatial units, as well as the methods used to analyse their internal composition of economic activities.

3.2.1 Study area

Belo Horizonte is the urban core of the third largest metropolitan area in Brazil, and is located in the heart of the country's iron ore mining region - the Iron Quadrangle (see Figure 3.2.1 and 3.2.2). The city is the capital of Minas Gerais, a state historically dominated economically by the mining industry. In recent years, after two mining accidents totalling almost US$ 19 billion in economic losses and almost 300 fatal victims (Sapata Gonzalez et al., 2022), not to mention an unpriceable environmental damage, growing discontent questioned the dominance of one single extraction industry in the state, with civil society pressure reaching transnational contexts (Cezne, 2019). In addition to being prone to deadly accidents, mining industry is expected to have a limited time-span, with mines reaching exhaustion and finally closing after a couple of decades. Dependency on mining also increases a place's vulnerability to external shocks, affecting labour market structures, households' income, and local governments' fiscal situations (Silva et al., 2021).

Knowing that the mining industry is still growing, with mineral extraction expected to increase threefold to fourfold within the next decade (Coelho et al., 2020), understanding the current structure of Belo Horizonte's economy will be helpful to plan for its future. Possible transitions towards a post-extractivist economy should include diversification strategies, human capital formation policies, and further public investment to increase local economic resilience (Silva et al., 2021). Reducing local dependency on mining can make use of complexification strategies at the local level, enabled by holistic environmental and social governance, already included in post-disaster mitigation efforts (Milanez et al., 2021) and in Brazilian legal framework in a broader sense (Haddad, 2015).

3.2.2 Data collection and filtering

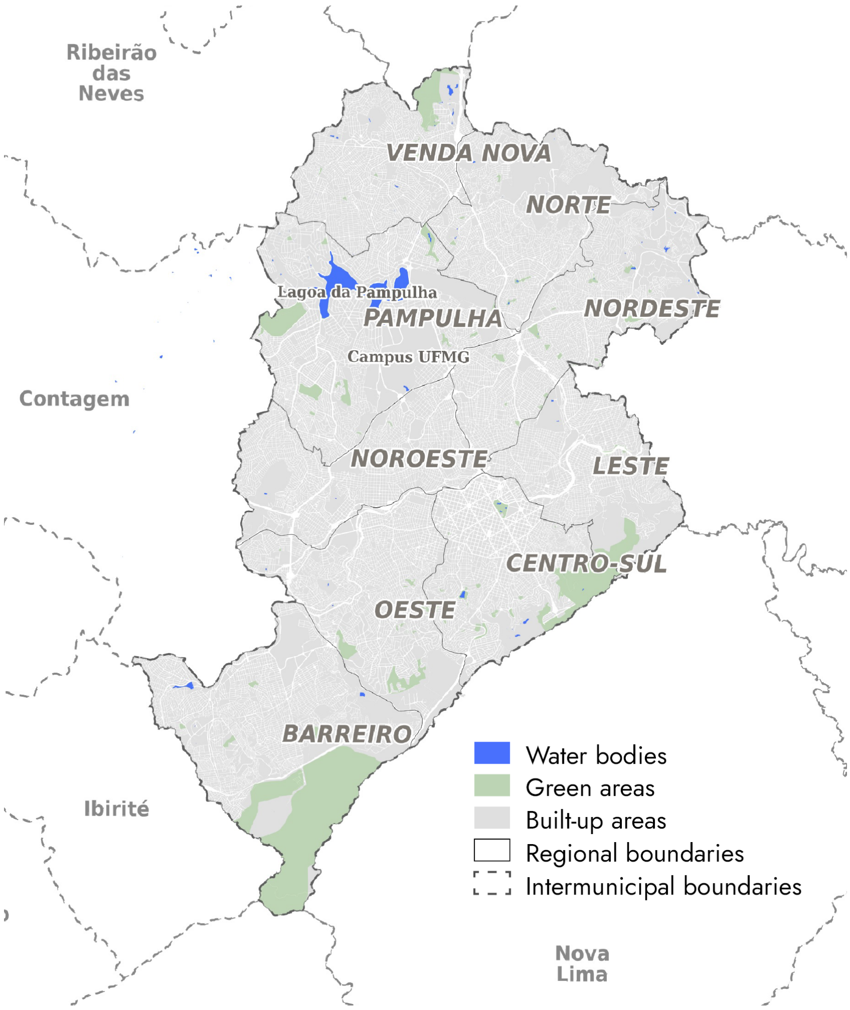
The main dataset used for this research is a Vector – Point dataset with all economic activities registered in the municipality. Each point refers to a single economic activity, characterised by its official names, addresses, area used within a building, the type of activity, among others. The categorisation of the type of activities follows a nationally standardised coding system called CNAE (*Classificação Nacional de Atividades Econômicas* - National Classification of Economic Activities in Por-

Figure 3.2.1: Location of Belo Horizonte in Brazil.

tuguese). This has been used in previous studies analysing the economic structure of Brazilian cities (Abreu et al., 2020; Barufi, 2018; Maraschin & Krafta, 2013). It is the Brazilian equivalent of similar coding systems worldwide, such as the Netherlands' SBI (*Standaard Bedrijfsidentificatie* - Standard Business Identification in Dutch) (Smit et al., 2015) and the U.S.'s North American Industrial Classification System (NAICS) (Sprague, 2018).

The dataset for the location of economic activities was available for the years of 2011, 2015 and 2019, allowing for temporal comparison throughout these 8 years. A thorough inspection of the mapped points did not detect mistakenly duplicate entries, mispositioned points or other inconsistencies. However, some degree of filtering was necessary for the purpose of the research. Three criteria were used: (1) whether economic activity's work is conducted within the registered and depicted address; (2) whether transaction or provider-customer contact is conducted at the depicted location; and (3) whether a location is characterised mainly by the presence

Figure 3.2.2: Map of Belo Horizonte, highlighting intermunicipal and intra-municipal (regional) boundaries.



of people (employers, employees, clients or others) in face-to-face contact (Storper & Venables, 2004) instead of machinery or goods. The filtering process included activities fulfilling at least one of these criteria and resulted in 199,407 entries for the year of 2019, 167,274 entries for the year of 2015 and 100,738 entries for the year of 2011.[1]

3.2.3 Spatial unit definition

The first step to define spatial units for the case study was to detect clusters of economic agglomeration by applying the Accessibility Index (Hidalgo & Castañer, 2015) for the filtered dataset of the most recent year (2019). The Accessibility

[1] It is important to highlight, at this point, that informal activities compose a high parcel of the Brazilian economy and are not included in this study due to the lack of data. Still, it is not believed to have influenced the results due to recent attempts by local and national governments to formalise individual entrepreneurs (e.g. domestic cooks, craftspersons, artists, etc.), who can become their own individual companies. Several entities of the sort were observed in the data.

Index (equation 3.1) helped deriving the basic spatial units for this research. To be able to compare different years, we defined spatial units for the most recent year available and applied their boundaries to the datasets for the previous two years. This was done to avoid the usage of pre-determined administrative spatial units that aggregate characteristics to arbitrarily defined limits. This way, the concentration of economic activities by itself generates this study's spatial-analytical units.

Following Araldi and Fusco's (2019) multi-scalar approach for analysing urban services, three levels were defined: Local, Neighbourhood and Community. Equation 3.1 describes accessibility A of activity i, based on all other actvities j, the distance d between i and j, and the constant γ used to detect three levels of clustering (Local, Neighbourhood and Community) by varying between 32, 16 and 8. For $\gamma = 32$, the influence of an activity (j) on another's (i) accessibility (A) decreases by half roughly every 21m, being neglectable at around 150m, defined here as the Local level of clustering. For $\gamma = 16$, the influence of an activity on another's accessibility decreases by half roughly every 42m, being neglectable at around 300m, defined here as the Neighbourhood level. For $\gamma = 8$, the influence of an activity on another's accessibility decreases by half roughly every 84m, being neglectable at around 600m, defined here as the Community level.

$$A_i = \sum_{j=1}^{N} e^{-\gamma d_{ij}} \tag{3.1}$$

For each level of aggregation, the accessibility values per activity are interpolated via an Inverse Distance Weighted (IDW) method, using an exponential function to weigh values by distance. The segmentation of the interpolated surfaces into units of analysis is conducted using SAGA's region-growing algorithm of Watershed Segmentation (Conrad et al., 2015). The algorithm selects local maxima as seeds, sectioning the area using the valley lines – i.e., areas with less concentration of economic activities – as edges, assigning for each spatial unit a unique peak of concentration to which it belongs. The result is a segmentation of the area into mutually exclusive spatial units polarised by points of high concentration of activities. To avoid oversegmentation, a peak-to-valley threshold equal to the raster's Standard Deviation was defined after a trial-and-error approach in line with the literature (e.g., Liu et al., 2018).

3.2.4 The Economic Complexity Index

The ECI builds on two metrics: an area's Balassa Index (BI, Hidalgo and Hausmann, 2009; Hidalgo et al., 2007) and the Revealed Comparative Advantage (RCA). The BI is calculated by detecting whether a region has a higher share of a certain service than average (equation 3.2). Here, the first differentiation of the usual application of the ECI: commonly, the BI is applied in trade flow analyses, detecting whether a certain product is more intensely exported by a region than average. The BI informs a binary RCA (equation 3.3). By applying this to all economic sectors of the city, an area of a city having a higher number of a certain business category than the average of all areas ($BI > 1$) indicates that said area has an RCA in performing that category of service or retail. The intuition behind it is the following: if an area has a high concentration of, say, car dealerships, it is an indication that some specialisation has taken place, with the right set of capabilities in place for it to happen. From the point of view of trade theory, an analogy is also possible: higher concentrations of psychological services, for instance, is an indication that such area *exports* this service to other areas of the city, since an inflow of consumers is expected for that product or service from elsewhere in the city. Equation 3.2 contains a business category identifier (k), a spatial unit identifier (i), and the total number of areas (K). The number of firms (a) belonging to business category (k) present in spatial unit (i) is represented here as $a(k,i)$. A derived measurement used is the total number of firms for one specific business category in all areas, represented as a_i.

$$BI_{k,i} = \frac{a(k,i)}{a_i/K} \tag{3.2}$$

$$RCA_{k,i} = \begin{cases} 0, & \text{for } BI_{k,i} < 1 \\ 1, & \text{for } BI_{k,i} \geq 1 \end{cases} \tag{3.3}$$

By positioning the binary values for RCA in a proximity matrix with rows as spatial units and columns as business categories, the ECI calculation derives the diversity of an area as the number of business categories in which it has a comparative advantage, and the ubiquity of a category as the number of areas with a comparative advantage in it. Hidalgo and Hausmann (2009) define the calculation of the index via a so-called *method of reflections*: the previously calculated diversity of an area is updated according to its categories' ubiquities; the previous ubiquity of a category is updated according to their areas' updated diversities; the diversity is again updated according to the average diversity of the other areas with comparative advantage in the same categories; the ubiquity of a category is again updated according to the average ubiquity of the other categories within the same areas; and so forth, in an iterative process. A satisfying level of iteration (i.e., convergence) is also met by considering the eigenvector $\vec{C}$ with the second largest eigenvalue of the proximity matrix. In the equation below, $\langle\vec{C}\rangle$ represents the average and $std(\vec{C})$ represents the standard deviation. Economic Complexity Index (ECI – Equation 3.4) for all areas k is then defined as follows:

$$ECI = \frac{\vec{C} - \langle\vec{C}\rangle}{std(\vec{C})} \tag{3.4}$$

BI, RCA and ECI were calculated using the R packages EconGeo (Balland, 2017) and EconomicComplexity (Vargas et al., 2020). Other derived products from these calculations are a complexity index for business categories assessing which categories were classified as complex, and proximity matrices for business categories, relating which categories tend to appear near one another. These are the base for generating the Product-Space network (Hidalgo et al., 2007), described in the following paragraph.

By using proximity matrices for business categories, it is possible to develop a network of categories based on the likelihood of them appearing within the same spatial unit, the Product Space (Hidalgo et al., 2007). The Product-Space gives meaningful insights on which business categories the ECI calculations are assigning as more or less complex. Its network form, with categories as nodes and probability of co-occurrence as edges, highlights how central nodes of economic activities are in relation to a whole network, indicating specific economic activities that enhance local complexity or fosters local diversification. This was detected by calculating the betweenness of this network's nodes.

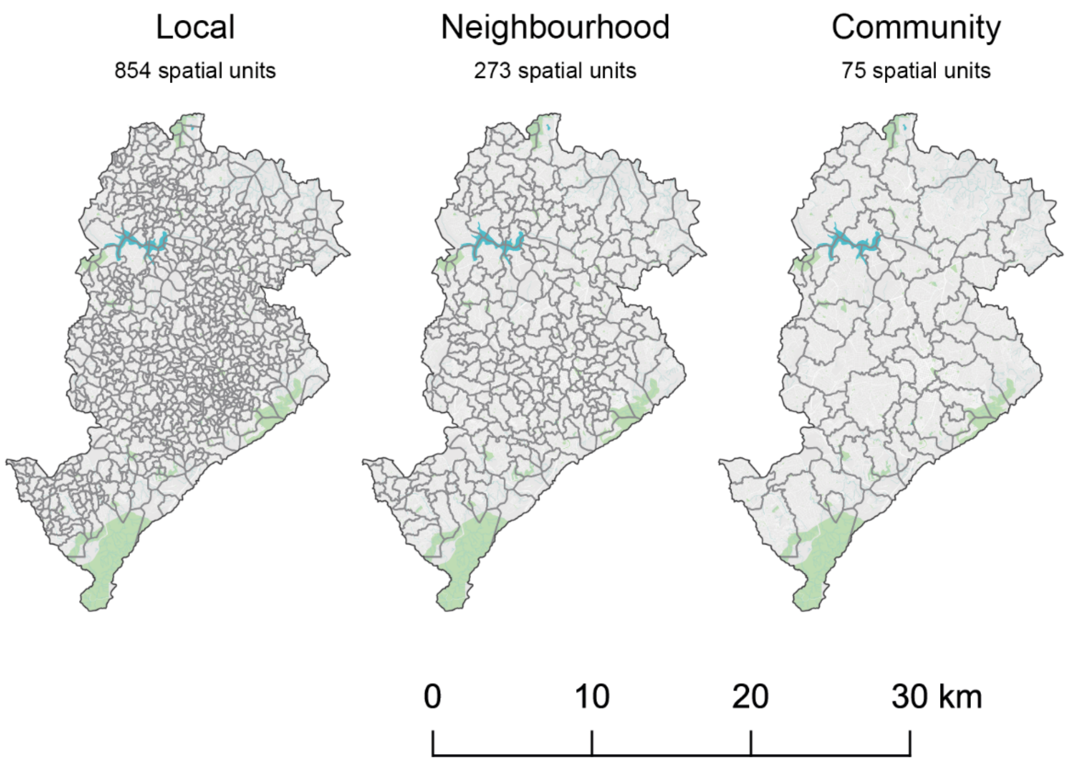
A network of Product-Space was built for the lowest level of aggregation (Local level) with the highest number of spatial units, for the most recent year of analysis (2019). The resulting graph was analysed using igraph (Csardi & Nepusz, 2006) and characterised in relation to the business categories themselves (2-digit level - Figure

3.3.4), the betweenness of nodes in the network, and the type of business categories assigned to the highest or lowest complexity values. Since this research's holistic usage of the Product-Space (Cicerone et al., 2019; Hidalgo et al., 2007) innovatively includes services in addition to manufacturing industries and amenities, considering the comparative advantage that business categories have by being more present than the average in each spatial unit, it was decided to give it the name of Urban Economy Space.

3.3 Results

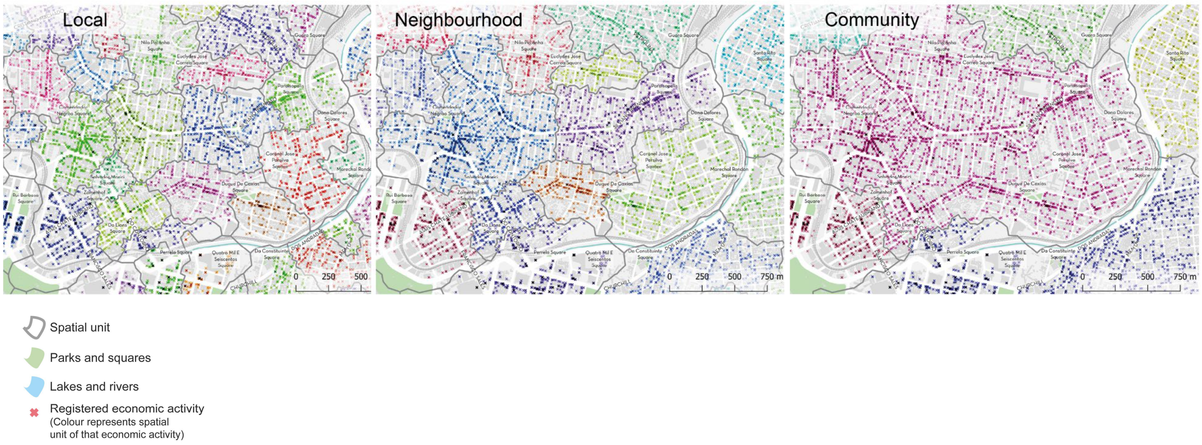
The bottom-up, data-driven definition of spatial units generated 854 units for the Local level (i.e., the lowest level of clustering), 273 units for the Neighbourhood level, and 75 units for the Community level (Figure 3.3.1).

Figure 3.3.1: Segmentation results defining the spatial units of this research for Belo Horizonte, Brazil.



Mapping the economic activities, detecting their concentration for different clustering levels, and segmenting the economic landscape into spatial units of analysis provide us already with interesting observations on how economic activities cluster in space. The Local level cluster detection for this research generated spatial units very close to the streetscape, being detected that the segmentations followed barriers also found in the landscape, such as larger street crossings, rivers, parks, among other areas with a lower concentration of economic activities. The same pattern is found to be reproduced for higher-level clustering. The Neighbourhood-level cluster detection is close in scale to the city's division of neighbourhoods, coinciding also with Regional boundaries when they exist. This can be seen in Figure 3.3.2.

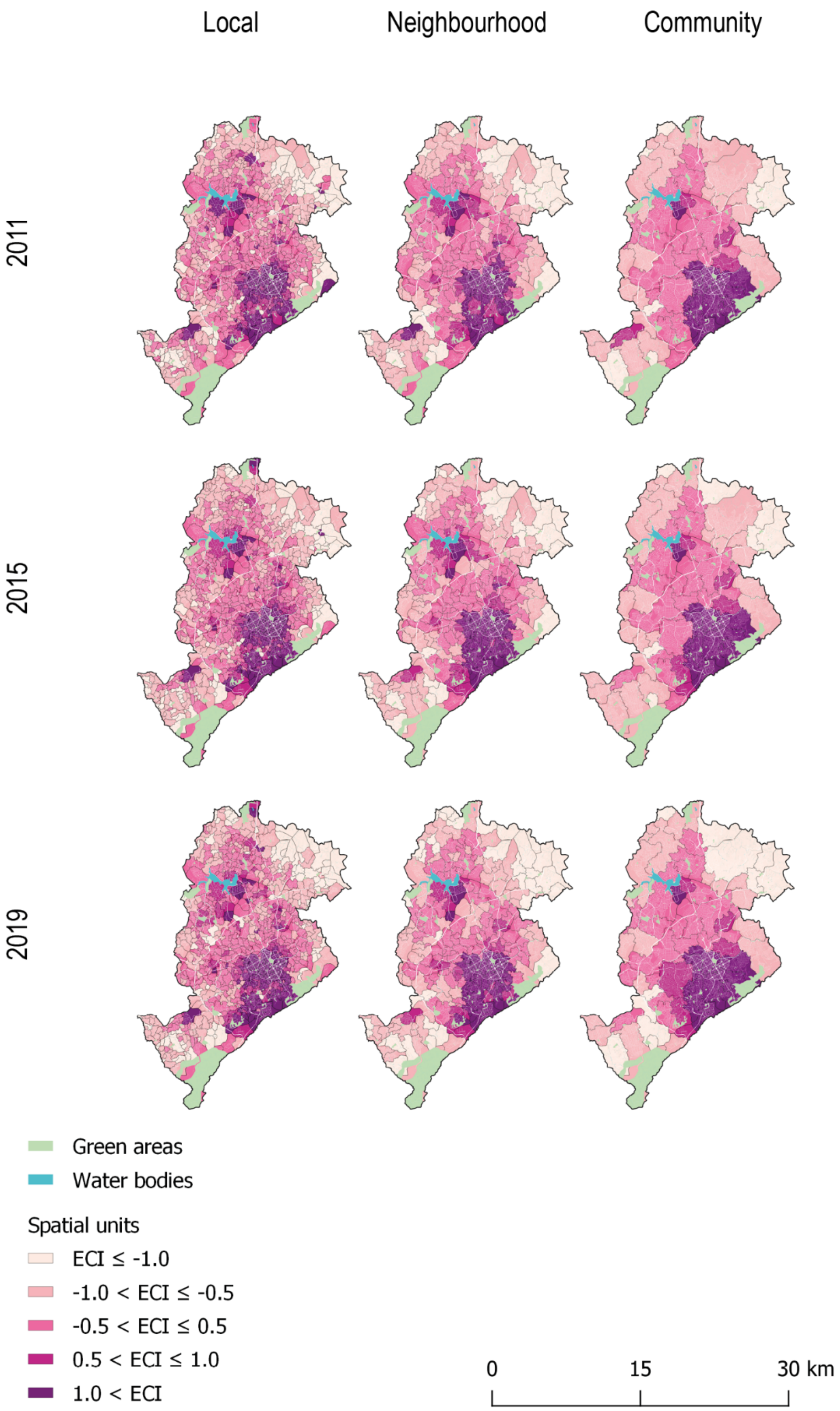
Figure 3.3.2: Close-up view of Santa Tereza neighbourhood, divided into the three levels of clustering used in this research.



3.3.1 Economic structure of spatial units

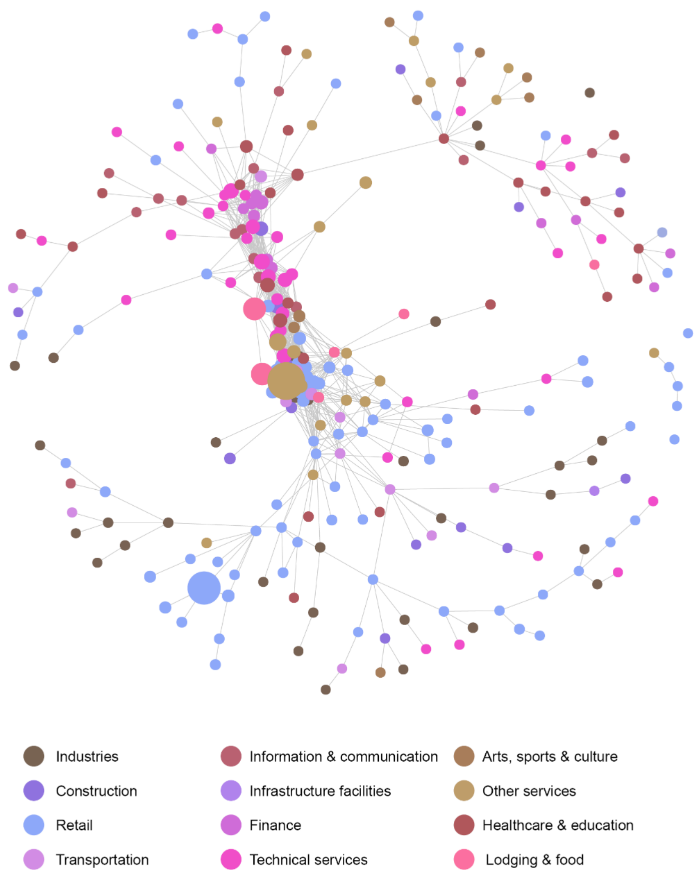
Figure 3.3.3 shows ECI distributed in Belo Horizonte across scales and over time. In order to be able to compare different years, the chosen approach was to define the clusters for the most recent year available and apply their boundaries to the datasets for the previous two years. It can be seen that the highest ECI values are predominantly located in the original urban core of the city of Belo Horizonte across all scales. This is an area with orthogonal street layout, planned in the 1890s, and nowadays constitutes the wealthiest and most densified area of the city. Local level clustering also detected local commerce centres with higher ECI values, as well as the spatial unit comprising Minas Gerais state's Administrative Headquarters, in the northernmost corner of Belo Horizonte – especially for years 2015 and 2019, after it was completed. The spatial units around the Federal University's campus also showed higher ECI values.

Figure 3.3.3: ECI per spatial units in the different years and spatial scales.



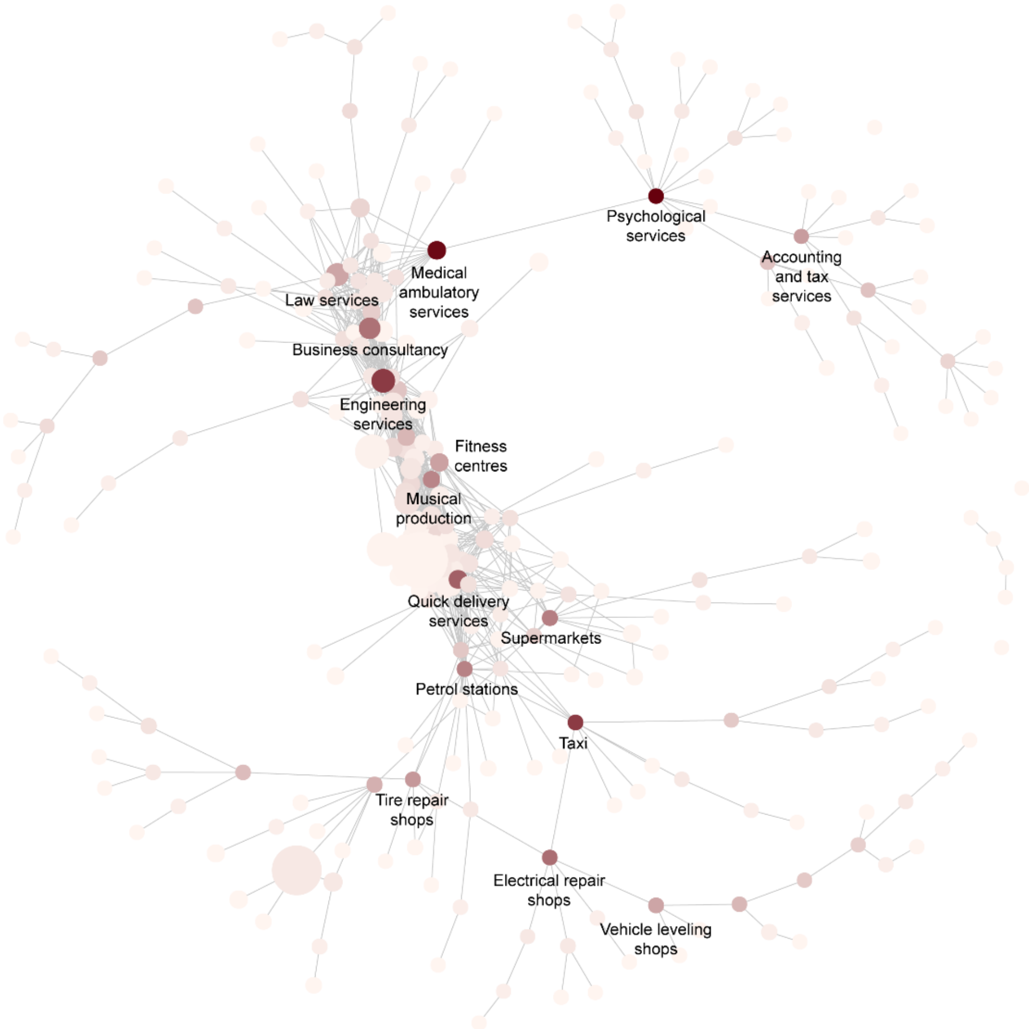
It is also important to understand how the ECI is related to the categorical composition of economic activities within these areas. The Urban Economy Space expands from the original Product-Space (Hidalgo et al., 2007) and the Amenity Space (Hidalgo & Castañer, 2015; Hidalgo et al., 2020) by considering all types of business categories, in place of exported products and in addition to neighbourhood amenities. By plotting these categories' co-occurrence in spatial units into a network, Figure 3.3.4 shows a remarkable central core concentrating most of the economic categories observed. The branches located in the bottom corner of the chart, for instance, are more dedicated to retail activities.

Figure 3.3.4: The Urban Economy Space of the city of Belo Horizonte, categorised by 2-level CNAE aggregation of business categories. Circle sizes indicate number of firms in each business category.



The strength of the connection between these nodes is determined by the relevance of co-occurrence between two business categories in each spatial unit. This indicates that there are categories that tend to co-occur in space more often than others, generating nodes pervaded by many lines. The calculated betweenness for these nodes (Figure 3.3.5) shows the business categories that tend to appear along most other business categories. These are not necessarily more complex, but their presence in a spatial unit indicate a more diversified economic landscape. These are detected to be pervasive categories of services either directed at households (e.g., fitness centres, vehicle repairing), or other businesses (e.g., accounting firms, consultancy).

Figure 3.3.5: Urban Economy Space classified by the betweenness of its nodes. Business categories with the highest betweenness are highlighted.



The ECI calculations assign complexity values for both areas of analysis and business categories. To further investigate which areas are considered complex, the complexity of business categories is assessed. More (less) complex business categories are expected to require higher (lower) capabilities, higher (lower) technological requirements, or more (less) knowledge-intensive firms. Plotting the business categories with highest (Figure 3.3.6) and lowest (Figure 3.3.7) complexity values associated with them confirmed these expectations.

Figure 3.3.6: Urban Economy Space with highest ECI business categories.

3.4 Discussion

3.4.1 The Urban Economy Space and ECI's transferability of scope

Firstly, the Urban Economy Space (Figure 3.3.4) shows a network with a remarkable core of activities, indicating activities that tend to appear together in the spatial units. Contrary to the expectations (Hidalgo et al., 2007), this entangled network of firm category co-location does not necessarily translate into an enhanced economic complexity, or sophistication. The complexity of the business categories in this core ranges in mid-values, connecting branches of higher complexity in the top-right corner (Figure 3.3.6) of the chart with branches of lower complexity in the bottom

Figure 3.3.7: Urban Economy Space with lowest ECI business categories.

corner (Figure 3.3.7). There are sub-networks of higher complexity and sub-networks of lower complexity, depending on the nature of economic activities that happen, and they both locate in the peripheries of the network.

Secondly, the analysis of betweenness in the network highlights that more generalised business categories act as catalysers for the presence of a higher number of categories and, therefore, a higher variety of activities. It can be concluded that the presence of categories such as law firms, consultancy firms, engineering services, supermarkets, fitness centres and music production, either fosters a richer composition of business categories within an area or is a consequence of it. In other words, the presence of these categories of firms is an indication of local economic diversity. This is an interesting finding since these categories do not appear to be themselves specialised. In fact, their level of generality – a law firm can act as auxiliary to a multitude of other business categories, for instance – may actually be the reason why they appear as strong catalysers in the first place.

The sort of business categories figured with the highest complexity values (Figure 3.3.6) is in line with categories highlighted by Murdoch, 2018 as demanding highly-skilled labour in the post-industrial economy and seen as the main sources of the economic prosperity for global cities (Sassen, 2005). These are financial, telecommunications, scientific and technical services, educational services, healthcare and social assistance, among others. In this intra-urban analysis, these categories were also seen to contribute to overall higher levels of complexity of the economy. It is important to highlight that some of these categories are characterised by having a strong or exclusive governmental presence in the Brazilian context, especially medical services, higher educational facilities, hospitals, and public administration services (government-run by definition). This indicates that local and national governments can have a role in increasing the complexity of local economies.

On the other side of the spectrum, Murdoch (2018) describes less advanced industry categories as being economically traditional: construction, manufacturing, retail trade, transportation and warehousing, among others. These coincide with the lowest complexity categories observed in the Urban Economy Space (Figure 3.3.7). This indicates a successful broadening in the scope when transferring applications of the Economic Complexity Index from the common country or regional level down to city and intra-city levels. Higher levels of aggregation (e.g. country-level), however, tend to position manufacturing industries within the higher complexity ones (Hidalgo, 2015), as opposed to commodity extraction and agriculture as low complexity. By transferring it to the city-scale, incorporating the service industry to the analysis, manufacturing figures among the lowest levels of complexity, possibly because agricultural and mining activities are even less commonly found in urban areas.

It is important to highlight, at this point, the nature of the calculation of the economic complexity index and how activities are assigned their own values of complexity. Since it is based on an imbalance by definition, as made explicit in the “method of reflections”, when applying it to an intra-urban area some specific activities can be highlighted as complex, although not being considered in other situations. That was the case, specifically, for parking lots. Our guess is that some specific parts of the city, with a specific aspect of the urban form, are highlighted as complex, leading along the economic sectors that are part of them. Central, richer neighbourhoods were continuously highlighted as highly complex. Those central neighbourhoods, for reasons not necessarily isolated by this research, attract inhabitants from the whole Metropolitan area, and this might be reason enough for them to filled with parking lots, for instance. These richer neighbourhoods may, again for reasons not isolated by this research, attract more close attention from the public sector and be more often chosen for hosting a variety of public services, such as schools and clinics. This does not exclude the possibility of local governments complexifying local economies, but rather poses a question to the use of the method itself. However, as the index is created by a pre-defined imbalance, so are our cities, so is our society, and so are different countries. This, perhaps, is even more reason to dedicate attention to these issues, and the overall unequal nature of the economic system that shapes our environment and, by extension, our studies.

3.4.2 Research directions

The nature of the dataset was an important limitation for this research. Since locational data for firms is dependent on a formal registration with municipal authorities, informal economic activities, very relevant in the context of Brazil, are not depicted by the dataset. This is especially relevant for poorer areas, specifically the slum areas (favelas) and the poor peripheries typical for Brazilian cities. This means that the ECI values calculated for these areas are most likely lower than the actual complexity of the place, i.e. in case informal activities were considered. Further research in the field are encouraged to try to map and include informal activities.

It is also encouraged for further research to make a more explicit connection between common scale of applications of the ECI (i.e., regional and national scales) to the urban and intra-urban scales addressed in this paper. Similarly, the connections between processes or services detected at a local scale and exported products at a national scale can be made more explicit in further research. Moreover, this paper considers the mere location of activities in space as indication of local specialisation and, as such, the object of analysis. Given that capabilities and production processes are bounded geographically (Hidalgo, 2015), exploring local drivers of economic complexity, such as urban morphology indicators, is a natural development of this research.

Using higher granularity for spatial units, or even a spatial continuum of measurements in place of spatial units at all, could allow for a multi-level approach that prevents a significant Modifiable Areal Unit Problem (MAUP, Gotway and Young, 2002). In general, studies related to complex systems and emergent patterns should try to model individual elements and their relations to surrounding as close in scale as possible to the actual elements and their characteristics. That leads to a further recommendation for future studies that is to explore possible theoretical and microecnomic foundations for the complexity-enhancing process of recurrent diversification and specialisation, making use of simulations and agent-based models, for instance.

3.5 Conclusion

This research highlights that local or even national governments can have an important role in complexifying local economies. Some of the business categories classified as complex commonly belong to the public sphere, such as higher education facilities, hospitals, and public administration services. It is valuable that the positioning of these facilities in cities takes the underlying objective of fostering local economic complexity into account. Besides acting directly, governments could also promote the enhancement of specific industry categories considered to be more complex, the enhancement of categories that are shown to be more connected to others (enhancing diversity), or, ideally, prioritising multiple business categories that act together as networks of enhanced complexity. These are concluded by this research as potentially more efficient manners of promoting economic performance than the common directed specialisation policies, usually aiming at single sectors (Hong & Xiao, 2016). Overall, this research has made use of innovative methods and has contributed to enhance current literature in the field of economic complexity. The application of the Economic Complexity Index to intra-urban areas, including the service industry, bridges a gap between the economics of development and urban and regional sciences.

References

Abreu, M. P., Del-Vecchio, R. R., & Grassi, R. (2020). Analysis of productive structure applying network theory: The Brazilian case. *Structural Change and Economic Dynamics*, *53*, 281–291.

Adler, P., Florida, R., King, K., & Mellander, C. (2019). The city and high-tech startups: The spatial organization of Schumpeterian entrepreneurship. *Cities*, *87*, 121–130.

Araldi, A., & Fusco, G. (2019). Retail Fabric Assessment: Describing retail patterns within urban space. *Cities*, *85*, 51–62.

Balland, P.-A. (2017). *Economic geography in R: Introduction to the EconGeo package*. Utrecht.

Balland, P.-A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., & Rigby, D. (2022). The new paradigm of economic complexity. *Research Policy*, *51*(3), 104450.

Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P. A., Rigby, D. L., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, *4*(3), 248–254.

Balland, P.-A., & Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, *93*(1), 1–23.

Barufi, A. M. B. (2018). Services that add value in the city: The rise of the modern economy in Brazil. *Cities*, *78*, 39–51.

Batty, M. (2017). Diverse cities, successful cities. *Nature Human Behaviour*, *1*(22), 2.

Bertani, F., Ponta, L., Raberto, M., Teglio, A., & Cincotti, S. (2021). The complexity of the intangible digital economy: An agent-based model. *Journal of Business Research*, *129*, 527–540.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, *104*(17), 7301–7306.

Bishop, A., & Mateos-Garcia, J. (2019). Exploring the link between economic complexity and emergent economic activities. *National Institute Economic Review*, *249*(1), R47–R58.

Bond-Smith, S. C., & McCann, P. (2020). A multi-sector model of relatedness, growth and industry clustering. *Journal of Economic Geography*, *20*(5), 1145–1163.

Boschma, R. A., & Fritsch, M. (2007, September 17). *Creative Class and Regional Growth Empirical Evidence from Eight European Countries* (SSRN Scholarly Paper No. ID 1023869). Social Science Research Network. Rochester, NY.

Burlina, C., & Antonietti, R. (2020). *From variety to economic complexity: Empirical evidence from Italian regions*. Utrecht.

Cezne, E. (2019). Forging transnational ties from below: Challenging the Brazilian mining giant Vale S.A. across the South Atlantic. *The Extractive Industries and Society*, *6*(4), 1174–1183.

Cicerone, G., McCann, P., & Venhorst, V. A. (2019). Promoting regional growth and innovation: Relatedness, revealed comparative advantage and the product space. *Journal of Economic Geography*.

Coelho, P. S., Cordeiro, M. P., & Massola, G. M. (2020). Effects of female political participation on insecurity in a mining area. *International Perspectives in Psychology: Research, Practice, Consultation*, *9*(1), 18–35.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Sy*, 1695.

Dosi, G., Mathew, N., & Pugliese, E. (2022). What a firm produces matters: Processes of diversification, coherence and performances of Indian manufacturing firms. *Research Policy*, *51*(8), 104152.

Faggio, G., Silva, O., & Strange, W. C. (2020). Tales of the city: What do agglomeration cases tell us about agglomeration in general? *Journal of Economic Geography*, *20*(5), 1117–1143.

Florida, R., Adler, P., & Mellander, C. (2017). The city as innovation machine. *Regional Studies*, *51*(1), 86–96.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5), 685–697.

Gala, P., Rocha, I., & Magacho, G. (2018). The structuralist revenge: Economic complexity as an important dimension to evaluate growth and development. *Brazilian Journal of Political Economy*, *38*, 219–236.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632–648.

Ha, L. T., Dung, H. P., & Thanh, T. T. (2021). Economic complexity and shadow economy: A multi-dimensional analysis. *Economic Analysis and Policy*, *72*, 408–422.

Haddad, M. A. (2015). A framework for urban environmental planning in Brazil. *European Review of Latin American and Caribbean Studies / Revista Europea De Estudios Latinoamericanos y Del Caribe*, (99), 113–125.

Hidalgo, C. A. (2015). *Why information grows: The evolution of order, from atoms to economies*. Penguin Books.

Hidalgo, C. A., Castañer, E., & Sevtsuk, A. (2020). The amenity mix of urban neighborhoods. *Habitat International*.

Hidalgo, C. A., & Castañer, E. E. (2015). *The amenity space and the evolution of neighborhoods*.

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, *106*(26), 10570–10575.

Hidalgo, C. A., Klinger, B., Barabasi, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487.

Hong, S., & Xiao, Y. (2016). The influence of multiple specializations on economic performance in U.S. Metropolitan areas. *Sustainability (Switzerland)*, *8*(9), 1–16.

Jacobs, J. (1970). *The economy of cities* (1st ed.). Vintage Books.

Liu, L., Peng, Z., Wu, H., Jiao, H., Yu, Y., & Zhao, J. (2018). Fast identification of urban sprawl based on K-means clustering with population density and local spatial entropy. *Sustainability (Switzerland)*, *10*(8).

Lo Turco, A., & Maggioni, D. (2022). The knowledge and skill content of production complexity. *Research Policy*, *51*(8), 104059.

Maraschin, C., & Krafta, R. (2013). Growth dynamic of retail locations: A methodological approach using a logistic model. *Cybergeo*, *2013*.

Marco, R., Llano, C., & Pérez-Balsalobre, S. (2022). Economic complexity, environmental quality and income equality: A new trilemma for regions? *Applied Geography*, *139*, 102646.

Mealy, P., & Teytelboym, A. (2020). Economic complexity and the green economy. *Research Policy*, 103948.

Milanez, B., Ali, S. H., & Puppim de Oliveira, J. A. (2021). Mapping industrial disaster recovery: Lessons from mining dam failures in Brazil. *The Extractive Industries and Society*, *8*(2), 100900.

Murdoch, J. (2018). Specialized vs. diversified: The role of neighborhood economies in shrinking cities. *Cities*, *75*, 30–37.

Nguyen, C. P. (2022). Does economic complexity matter for the shadow economy? *Economic Analysis and Policy*, *73*, 210–227.

Poncet, S., & Starosta de Waldemar, F. (2013). Export Upgrading and Growth: The Prerequisite of Domestic Embeddedness. *World Development*, *51*, 104–118.

Romero, J. P., & Gramkow, C. (2021). Economic complexity and greenhouse gas emissions. *World Development*, *139*, 105317.

Salvati, L., Mavrakis, A., Serra, P., & Carlucci, M. (2015). Lost in translation, found in entropy: An exploratory data analysis of latent growth factors in a Mediterranean city (1960–2010). *Applied Geography*, *60*, 107–119.

Sapata Gonzalez, R., Aparecida da Silveira Rossi, R., & Gustavo Martins Vieira, L. (2022). Economic and financial consequences of process accidents in Brazil: Multiple case studies. *Engineering Failure Analysis*, *132*, 105934.

Sassen, S. (2005). The global city: Introducing a concept. *Brown Journal of World Affairs*, *11*(2), 27–43.

Sciarra, C., Chiarotti, G., Ridolfi, L., & Laio, F. (2020). Reconciling contrasting views on economic complexity. *Nature Communications*, *11*(1), 3352.

Silva, L. B., Comini, I. B., Alves, E. B. B. M., da Rocha, S. J. S. S., & Jacovine, L. A. G. (2021). Compensating the negative environmental impacts of mining with financial mechanisms in Brazil. *Land Use Policy*, *104*, 105351.

Smit, M. J., Abreu, M. A., & de Groot, H. L. (2015). Micro-evidence on the determinants of innovation in the Netherlands: The relative importance of absorptive capacity and agglomeration externalities. *Papers in Regional Science*, *94*(2), 249–272.

Sprague, C. D. (2018). *Resilience and the U.S. labor market: A cross-scale analysis on the role of industrial diversity and specialization*.

Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370.

Vargas, M., Bottai, C., Kozlowski, D., Pintar, N., Bank, T. W., & Statistics, O. T. (2020). Computational methods for economic complexity.

Xiao, Y., & Drucker, J. (2013). Does economic diversity enhance regional disaster resilience? *Journal of the American Planning Association*, *79*(2), 148–160.

Zhu, S., Yu, C., & He, C. (2020). Export structures, income inequality and urban-rural divide in China. *Applied Geography*, *115*, 102150.

Chapter 4

Digital or local? Impacts of online shopping and teleworking on urban expansion: a theoretical approach

The ongoing shift towards online shopping, a trend further accelerated by the Covid-19 pandemic, has impacted urban spaces. Coupled with the changes in working habits, centrally located retail and office spaces might be severely affected. Despite these disruptive trends, the potential drop in rental and property prices, resulting from decreased urban demand, could revive attraction towards city centres, generating conflicting forces that need to be further studied. In order to address these issues, this paper proposes a simplified linear city model. Departing from the classic monocentric city model, it aims to emulate the interactions between households and firms, contributing to the discussions around the future of consumption, retail spaces, and urban form. By examining the bid-rent curves generated by different consumption behaviours, the paper explores how the interplay between various factors such as online shopping behaviour, import costs, and frequency of shopping and commuting trips can shape the built-up landscape of a fictitious city. Analysis of the equilibrium suggests that different consumption behaviours might directly affect the viability of central retail and even disrupt urban fabric, indicating the importance of the resilience of local governments in the face of changes in working habits and consumer behaviours.

4.1 Introduction

Recent trends in online shopping have been intensified by the Covid-19 outbreak, with consumers moving from local shops and malls to internet-based orders. Even local shopping habits by excellence, such as groceries, have been impacted by these trends, with supermarkets offering delivery of goods to the doorstep. In some contexts, the increase was up to five-fold, with now more than 50% of survey respondents claiming buying online at least once a week (Young et al., 2022). The rise of personalised shopping experiences and mobile-optimised websites and apps for instance, together with advancements in logistics infrastructure, suggest that these are long-term trends, more related to changing attitudes and behaviours (Wieland, 2023a). These trends alone already generate an expectation of long-lasting impact on urban spaces, causing a downward pressure on demand for traditional centrally located retail floor-space, for instance, as already indicated by rising vacancies (Orr et al., 2023). Another already visible consequence is a surge in alternative spaces that support this new economy, such as the concept of 'dark kitchens' (Souza et al., 2022), warehouses and suburban logistic spaces.

Nonetheless, these concerning trends are also coupled with changes in the working culture (Barrero et al., 2021). It is evident that the pandemic has also accelerated previously existing trends in teleworking, further decreasing the demand for central offices and urban dwellings. However, these are shifts that can act both ways, and the long-run general equilibrium effects are not yet clear. For example, this urban flight can make central areas attractive again due to a downward pressure on rental and property prices (Florida et al., 2021). This can attract a certain parcel of the population back to central areas; a phenomenon that is not new and, sometimes, is also coupled with a change in character of traditional neighbourhoods which is often referred to as 'gentrification' (Glaeser et al., 2023).

The interaction of all these forces remains uncertain: it challenges the standard trade-off between housing and commuting costs, which commands the distribution of the population along the distance to the main job centres. Moreover, it questions the local push-pull factors that command clustering and subcentre formation, to where people make short-distance trips. It is vital to understand how these changes may impact locations of people and service-shops, their derived trips, and whether government reactions can improve sustainability and welfare.

This paper builds up a base urban model with households and firms in a theoretical monocentric city. It is used to contribute to discussions about the future of consumption, retail spaces, and urban form, with a focus on welfare and macro-patterns emerging from these interactions. The model encapsulates various factors such as online-shopping behaviour, importing costs, commuting and shopping trips, frequency of teleworking, a simplified rental market, and increasing returns to scale for firms. More specifically, the objective is to understand the impact of online shopping and teleworking on the urban form and viability of retail. We aim to answer three research questions:

- Which urban forms emerge from differing consumption preferences between local shopping and online shopping?
- What are the consumption preference conditions under which central retail is economically viable?
- How does an external shock of teleworking impact the resulting land-use patterns in this theoretical city?

We start the paper with a short overview of literature (section 4.2), highlighting studies connecting e-retailing to traditional retail, the impacts of teleworking to the urban form, and modelling approaches ranging from urban economics to retail location modelling. Secondly, we define the base model, inspired by the monocentric city literature (section 4.3), Thirdly, we derive the equilibrium of the model, highlighting relevant comparative statics to answer the three research questions (section 4.4). Finally, we conclude by highlighting the main discussions and how they can be expanded into further work (section 4.5).

4.2 Background

In this section, we first (section 4.2.1) shortly review how the literature tackles the challenges of e-shopping and changes in retail in order to position the model and the scope of the modelled agents. We then remind (section 4.2.2) the urban economic frameworks used to discuss how interactions between firms and households are adapted to retail.

4.2.1 Emerging dynamics in the retail landscape

Retailing can be defined as an exchange of physical goods (Anderson et al., 2003). This puts the geographical dimension in the centre of the subject, as either the good must be transported to the customer, or the customer must transport themselves to fetch it. Both of which are posed as a burden to either firm or customer, creating a disutility that tends to be avoided. Even before the rise of department stores and discount stores as we know them, customers tried to avoid the burden of traveling to shop via mail-order catalogue retailing, for instance (Anderson et al., 2003). As the main locus of consumption, the internal structure of cities is brought to the eye of the discussion, as access to diverse services and amenities is one of the main reasons why households agglomerate (Glaeser et al., 2001).

However, the burden of traveling was always there: shops agglomerating to provide a higher utility to customers. At the scale of the street, brick-and-mortar retailing tend to cluster in so-called 'high-streets', forming a series of local centres with differring levels of specialisation and hierarchy (Araldi & Fusco, 2019). At a more local scale, analyses have also been done on the internal structures of these high streets (Rao, 2019; Vaughan, 2015). However, these are not always only local retailers. Franchise branches, often multinational companies with pricing and location decision controlled from distant administrative headquarters, already compete for location at these high-streets and, more recently, the suburban shopping centre. Analysis of the internal structures and how people perceive the distribution of shops within a shopping centre has also been found in literature (Torrens, 2023). The shopping centre in itself can be seen as a two-fold attempt to overcome transportation disutilities: by both agglomerating stores and moving them closer to where customers are, in an increasingly suburbanising world.

Parallelly, we can also think of an increasingly digitalised world. The illusion of a frictionless consumption society, as predicted by many in the eve of widespread internet availability, has not come into place (Brynjolfsson & Smith, 2000). However, we have definitely seen a recurring process of disintermediation and reintermediation: reselling platforms, even via social media, have gained an impressive central role in recent years. Retail sectors have moved from single-channel, where only shopping in-person was possible, to multi-channel (both online and in-person), or even omni-channel, when the same firm offers the product parallelly to the client via different acquisition possibilities (Anderson et al., 2003). This has already caused significant industrial and transportation impacts, respectively referring to changes in location of facilities (e.g. warehouses), and delivery channels (e.g. shipping). One recurring question remains relevant: what impact does it have on traditional retail spaces?

In order to dwell briefly into this discussion, we must recall some closely related concepts. The effects of e-shopping on in-store shopping can be of four different natures: complementarity, substitution, modification or neutrality (Arranz-López et al., 2023). Respectively, they refer to whether e-shopping increases in-store shopping, decreases it, alters its dynamic, or does not affect it at all. Several studies have found evidences of all natures (see Arranz-López et al., 2023 for further references). Whether and to which extent e-shopping substitutes or complements traditional in-store shopping feed into two conflicting theories in the field: the information-diffusion theory vs. the efficiency theory (Anderson et al., 2003). The former claims that online shopping will be more prevalent among households who are more prone to the new platforms that allow them, i.e., higher-order cities more well connected to information and communication technologies (ICT). The latter claims that online shopping will be more prevalent among those who need it the most, by being less intensively served by a varied amalgam of stores, i.e., suburbs and rural areas.

Arranz-López et al. (2023) examine the impact of e-shopping on frequency and duration of shopping trips, finding complementarity in number of trips, but rather shorter shopping trips for those who shop online. Beckers et al. (2021) emphasise the challenges faced by local retailers with a surge in e-commerce, incorporating e-commerce into traditional retail location models in a subsequent work (Beckers et al., 2022). Farag et al. (2007) show empirically that online product searches can increase shopping trip frequencies, highlighting the complementarity effect. Other studies have shown that, beyond consumer characteristics, store features and accessibility play an important role in determining whether more shopping online or in-store will happen (Teixeira et al., 2022; Wieland, 2023a; Wieland, 2023b). More recently, the particular effect of the COVID-19 pandemic was considered, highlighting its potential long-lasting impacts on local shops, adding to an already problematic decreasing share of retail-spending among households (Wieland, 2023b).

Not only e-commerce can threaten traditional retail space: the shift towards remote work, especially for high-income business service workers previously allocated in traditional employment centres, threaten the viability of retail and local consumption amenities (Althoff et al., 2022). More often than not, these two forces may feed on each other, with remote workers consuming more online, and vice-versa, a pattern that has also been exacerbated since the COVID-19 pandemic (Florida et al., 2021). This puts extra pressure on retail establishments, especially in traditional centres. These changes can be mirrored by the housing markets, for instance, as workers move to more affordable urban areas (Brueckner et al., 2023). In parallel, it was observed a more intense flattening of the rent gradient curve (Gupta et al., 2021), but also a more significant increase of rents in areas where telework was predominant (Gamber et al., 2023). Even within cities this could be noticed, with a downward pressure on rents in city centres with an increase in suburban areas (Gupta et al., 2021). While broader network relations between cities, globally, might remain unaltered, this phenomenon challenges traditional forces within cities, posing changes to office, retail, and housing floor-space availability and distribution. It is still worth mentioning that the conversion between these different uses is not an easy task, and demand high investments by both public and private actors (Gupta et al., 2023). Urban economic models emerge as essential tools to analyse this changing landscape of intra-urban structures and forces.

4.2.2 Modelling approach

Based on the discussions on the transformation of urban landscapes, it is crucial to opt for an analytical framework that can capture these complexities. Accordingly, we choose the lens of the monocentric city model. Introduced by Alonso (1964), it consists of a simple linear model with microeconomic foundations. In its market equilibrium, households in the same city are equally satisfied by consuming less of the housing good as they locate closer to an exogenous city centre. This happens since they pay higher rent per unit area, which is offset by a lower total transportation cost. Lower housing consumption on the same unit of land generates a rising density gradient towards the unique city centre. This model has been further expanded by

Muth (1969) and Mills (1972), onto what has become known as the Alonso-Müth-Mills (AMM) model. The work of Alonso, Müth and Mills builds on the classical von Thünen's (1863) Isolated State model and its implications for agricultural land-use. The set of rents producers are willing to pay as they move further away from a central location was hereafter referred to as the "bid-rent curve" and based on most studies in the field *a posteriori*. This includes Alonso-Müth-Mills models which analogously applied the concept to consumers instead of producers, and to housing rents instead of agricultural rents, in a monocentric city.

Retail has been explicitly considered in the urban economics framework by previous studies (Chu et al., 2012; Colombo & Hou, 2021; Guo & Lai, 2017; Lai, 2023; Ushchev et al., 2015). Historically, retail has been rather analysed along Huff (1966)'s probabilistic gravity-based approaches in a very empirical manner or through the spatial competition theories pioneered by Hotelling (1929). Chu et al. (2012) examine variety and price competition between firms without emphasizing the spatial dimension. Other studies, such as Colombo and Hou (2021) and Guo and Lai (2017) considered the spatial dynamics of retail competition between physical stores against an online competitor, but often lacking an explicit land market. This omission resulted in models not accurately reflecting the bid-rent theory's principles, evident when a baseline online consumption is treated as independent of location (Guo & Lai, 2017). Empirical studies such as Teulings et al. (2017) show, however, the relevance of the monocentric city model for analysing land markets around retail. In their case-study for the Netherlands, they show that shopping areas follow a land-market structure similar to that of a monocentric city, with rents decreasing around 15% for every 100m away from a centre (Teulings et al., 2017).

Expanding this scope, polycentric models initially introduced by Fujita and Ogawa (1982) and Ogawa and Fujita (1980) add an extra layer to the traditional monocentric approach, which may be particularly relevant for the debate between downtown shopping and emerging peripheral retail. A synthesis of these models with monopolistic competition, as attempted by Ushchev et al. (2015), offers valuable insights into firm relocation dynamics, while still not considering online shopping. The possibility for firms to relocate would dive us into the location competition discussion, yet we limit the scope of this paper to the competition between downtown and online shopping. With this, we can focus on the traditional trade-off between housing and accessibility to employment being challenged by the added dimension of a trade-off between costs associated with local shopping trips (local shoppers) and costs associated with online imports (online buyers).

4.3 The model

We define an urban economic model with retail including both online and downtown shopping derive the equilibrium and perform comparative statics (section 4.4). We describe the basic setting and the two types of agents (firms and households) in the next paragraphs.

4.3.1 Environment and basic settings

The environment consists of a one-dimensional (linear) landscape with discrete locations hereafter referred to as lots. The lots are initially occupied by farmers paying a uniform agricultural rent (R_A), assuming their products incur no transportation costs, and an exogenously defined Central Business District (CBD) in the origin. The CBD provides employment for households, which pay rent to absentee landowners. Households have the option of either consuming a uniform product sold by retail firms (hereafter referred to as shops) or importing the same product, and commute to the CBD. Shops pay rent, buy products from the outer-world by a fixed cost per product and a fixed importing fee, sell them for a profit to households. As a dynamic open city model, a new household enters the landscape until the long-run equilibrium condition is satisfied. Firms have free entry to the unlimited retail space of the CBD, as long as they can make positive (or zero) profit. Household and firm behaviours, interaction dynamics and parameter definitions are described in the next sections.

4.3.2 Household behaviour

Households in the model maximise their utility U by consuming land (S), and the uniform product sold by shops (Z). Utility is defined as a Cobb-Douglas function of these two variables:

$$U(S, Z) = S^{\alpha} Z^{\beta} \tag{4.1}$$

where $\alpha, \beta \in [0, 1]$ and $\alpha + \beta = 1$. This utility is subject to a budget constraint (Equation 4.2), where a household's exogenously defined disposable income equals their income Y less of transportation costs, and is fully distributed among rent and consumption of a shop's products as functions of the household's location x. The budget constraint function is, then, defined as:

$$Y - t(x) = r(x)S + Z(x)p \tag{4.2}$$

In this scenario, $t(x)$ represents the total transportation cost for a household located at x. The rent per unit of land paid at location x is given by $r(x)$, and p denotes the unit price for products. It is essential to mention that households can choose between two possible behaviors: they can either import the uniform product from abroad, incurring an import fee k, or purchase the product locally. By importing, they save on transportation costs and pay a lower price per product - the unit cost c. However, the import fee k is the same as the one paid by the shops themselves, which can benefit from economies of scale.

This concept emulates online shopping behavior, with its associated conveniences and disutilities. The absence of a shopping trip in a household's total transportation costs can be seen as savings not only in transportation but also in opportunity costs from shopping online. Conversely, the fixed import fee k accounts for the delivery costs of a product and can be interpreted as encompassing other disadvantages of online shopping (e.g., delivery waiting time). The online shopping behavior is specified in the model as the subscript e. By choosing this behavior, the household's budget constraint function is defined as follows:

$$Y - t_e(x) - k = r(x)S + Z(x)c \tag{4.3}$$

with their disposable income now net of importing fee k, and the unitary price for product Z as the cost c.

As the consumption exhausts the income of households, we can subject equation 4.1 to the constraining equations 4.2 and 4.3. Their first-order conditions yield the following utility-maximising Marshallian demands:

$$S^* = \frac{\alpha\,(Y - t(x))}{r(x)}, Z^* = \frac{\beta(Y - t(x))}{p} \tag{4.4}$$

for the local consumer, and:

$$S_e^* = \frac{\alpha(Y - t_e(x) - k)}{r(x)}, Z_e^* = \frac{\beta(Y - t_e(x) - k)}{c} \tag{4.5}$$

for the importing consumer.

By substituting the Marshallian demand functions defined in Equations 4.4 and 4.5 into the utility function defined in Equation 4.1 we reach the optimal utility functions $U^*(S^*, Z^*)$ and $U_e^*(S_e^*, Z_e^*)$, leading to the indirect utility functions $\upsilon(x)$ and $\upsilon_e(x)$ with utility depending on income and prices in relation to household location x.

$$\upsilon(x) = \left(\frac{\alpha(Y - t(x))}{r(x)}\right)^{\alpha} \left(\frac{\beta(Y - t(x))}{p}\right)^{\beta} \tag{4.6}$$

$$\upsilon_e(x) = \left(\frac{\alpha(Y - t_e(x) - k)}{r(x)}\right)^{\alpha} \left(\frac{\beta(Y - t_e(x) - k)}{c}\right)^{\beta} \tag{4.7}$$

It is worth mentioning that, as an open city model, for any state of the model there will be a net income of households as long as there is a location x providing a higher utility than a predefined outer-world utility (u_0). This also indicates the short-run equilibrium condition (4.8) for every state of the model: with freely mobile agents, the utility reached at the city's fringe ($\bar{x}$) will define a fixed utility $\bar{u}$ for every other household agent already present in that state.

$$\bar{u} = \max\left(\upsilon(\bar{x}), \upsilon_e(\bar{x})\right) \tag{4.8}$$

This short-run equilibrium condition allows to derive the bid-rent functions ($\Psi(x)$ and $\Psi_e(x)$) for every occupied location x depending on the fixed utility $\bar{u}$.

$$\Psi(x) = \bar{u}^{-1/\alpha}\alpha(Y - t(x))^{1/\alpha} \left(\frac{\beta}{p}\right)^{\beta/\alpha} \tag{4.9}$$

$$\Psi_e(x) = \bar{u}^{-1/\alpha}\alpha(Y - t_e(x) - k)^{1/\alpha} \left(\frac{\beta}{c}\right)^{\beta/\alpha} \tag{4.10}$$

Following the bid-rent theory, the land will be occupied by agents practising the behaviour that allows them to bid the highest for a specific location. The implications of this are discussed in Section 4.4.1. Without loss of generality, land will be occupied by the the highest-bidding behaviour between the two, as absentee landowners allocate land to the highest possible bidder, and by consequence the effectively paid rent is defined as follows:

$$r(x) = \max\left(\Psi(x), \Psi_e(x)\right) \tag{4.11}$$

4.3.3 Transportation cost function

It is now essential to discuss the transportation cost function. For simplicity, transportation costs are assumed to vary linearly with the distance traveled. All households are assumed to commute to the CBD for exchanging labour for income, while locally consuming households may incur additional trips to the CBD to shop. This differentiated behaviour with respect to travelling induces us to introduce another parameter, F, that represents the frequency of traveling to the CBD. For locally consuming households, F is a sum of their shopping and working trips, thus $F = F_S + F_W$. For a household that imports the product, their frequency of travelling to the CBD is limited by their working trips, F_W, thus $F_e = F_W$. For that reason, wihtout loss of generality, it can be said that $F \geq F_e$.

The "equal" part of that inequality refers to the possibility that a household does all its shopping in the same trip as their working trips. This would imply that $F = F_W$ and therefore $F = F_e$. However, it is not impossible to presume that a locally consuming household incurs in more trips than their usual working trips. There are multiple possible intuitions for this: households might prefer commuting to work by public transit and shopping by car; they might shop at entirely different times of the day; shops could be closed during their available time throughout the week; among others. All this leads us to both transportation cost functions, defined as follows:

$$t(x) = \tau F d(x) \tag{4.12}$$

$$t_e(x) = \tau F_e d(x) \tag{4.13}$$

where τ is the cost of travelling per unit of distance and $d(x)$ is the Euclidean distance between location x and CBD located at the origin.

4.3.4 Firm behaviour

A firm is assumed to be a monopolist, or part of an oligopoly as further firms enter the landscape. A firm j's revenue R_j (Equation 4.14) depends on quantity sold q_j. Similarly, its costs C_j (Equation 4.15) equal the quantity sold, multiplied by a unitary cost c for each product bought at an external market. A fixed importing fee (k), the same for both firms and households that decide to import the product by themselves, is added to a firm's cost. This introduces economies of scale to firms: the higher the quantity produced, lower will be the relative average cost of the marginal product. Finally, a firm has to pay rent equivalent to that of the highest bidder in the centre. It is assumed that a firm will always just marginally outbid a household for land. Therefore, for simplicity of notation, instead of referring to rents paid by firms as $r(x)$, we refer to it as $r(0)$: the rent effectively paid by firms will be the same the highest-bidding household would pay for the same location.

$$R_j(q_j) = pq_j \tag{4.14}$$

$$C_j(q_j) = cq_j + k + r(0) \tag{4.15}$$

For the context of this research, prices (p) are exogenously defined. Fixing prices exogenously is a strong assumption, but it can be often found in real life. For example, certain chain stores or retail franchises may not have the autonomy to set their own selling prices. They must accept prices for their products as given and still need to make decisions about shop locations based on market penetration and land prices. This scenario applies to fast-food chains and the fast-fashion industry, among others. In some cases, these shops might not even have central management or coordination, leading to spatial competition between local branches of the same brand. The fixed-price setting can also be interpreted as firms providing products with government-determined prices. This occurs in some contexts for liquor stores (e.g. Sweden) or petrol stations (e.g. Luxembourg).

There is an endogenously defined number of firms J in the city centre, dependent on parameter choices and utility level $\bar{u}$. All firms sell the same level of output q, a fraction of the total output Q divided by total number of firms J. We, thus, reach:

$$q_j = \frac{Q}{J} \tag{4.16}$$

4.4 Equilibrium and comparative statics

In this section, we first define all equilibrium conditions for the model. We then move to analyse the land-use patterns that arise in this equilibrium (section 4.4.1), the conditions of profitability for central retail (section 4.4.3) and the effects of a possible external shock of increasing teleworking (section 4.4.4).

Land market is cleared generating the first short-run equilibrium conditions for households, for each x:

$$r(x) = \max\left(\Psi(x), \Psi_e(x), R_A\right) \tag{4.17}$$

All households consume their optimum choice of housing and uniform product, choosing between the two possible consumption behaviour, whichever yields the highest utility, being the utility of all households in the landscape the same as of the last-entering one at the city's effective fringe $\bar{x}$:

$$\bar{u} \equiv \upsilon_{\bar{i}}(\bar{x}) = \max\left(\upsilon(\bar{x}), \upsilon_e(\bar{x})\right) \tag{4.18}$$

As an open-city, net migration will be zero when the utility at the fringe $\bar{u}$ equals that of the outer-world (u_0), generating the long-run equilibrium condition:

$$\bar{u} = u_0 \tag{4.19}$$

The total number of firms J will be such that each firm makes zero economic profit ($R_j - C_j = 0$). By combining that with Equation 4.16 we reach:

$$J = \frac{Q(p-c)}{k + r(0)} \tag{4.20}$$

And, finally, the total quantity of the output Q of firms will be the sum of all products locally consumed by all I households:

$$Q = \sum_i^I Z_i^* \tag{4.21}$$

4.4.1 Land-use patterns of consumption behaviour

Based on the two possible behaviours of households discussed and the equilibrium conditions, their respective bid-rent curves will determine which sort of behaviour will prevail and where this will happen. To analyse this, according to the input parameters and their comparative statics, is essential to understand what are the consequences of consumption behaviour to land-use patterns. Initially, we must discuss if the two bid-rent curves intersect at all in the positive quadrant and, if not, which behaviour will prevail according to the input parameters for all x. Secondly, if the two curves do intersect, two possibilities emerge: either locally consuming households inhabit near the city centre, or in the periphery. We explore, also, the conditions for each one of these situations to occur. Finally, there is a third possibility of an integrated landscape: one which both bid-rent curves coincide for all x, and under which consuming locally or importing the good is indifferent for all households.

For the two bid-rent curves to intersect in the positive quadrant ($\Psi(x) = \Psi_e(x)$), there must exist at least a point $\hat{x} \in R_+$ at which the equality holds. Conversely, if that point does not exist, it can be said that there is no intersection and either one of the behaviours prevails throughout the whole landscape. To test that, we first find the closed-form solution for point $\hat{x}$ (Equation 4.22). For simplicity, we consider the unitary transportation cost as the *numéraire* ($\tau = 1$).

$$\hat{x} = \frac{\left(\frac{p}{c}\right)^{\beta}(Y-k) - Y}{\left(\frac{p}{c}\right)^{\beta} F_e - F} \tag{4.22}$$

For the conditions of existence of the point $\hat{x}$, we will discuss the situations at which it does not exist in the positive quadrant. One of such situations is when the denominator of Equation 4.22 equals zero. This would mean $\left(\frac{p}{c}\right)^{\beta} = \frac{F}{F_e}$. In other words, when the price relation between local consumption and importing, weighted by the dedication of income to product Z (β), equals the relation between frequencies of city centre visits for both behaviours, the two bid-rent curves do not meet and only of one the behaviours prevails. In graphic terms, this situation generates an appearance of parallelism between the curves.

The other situation of non-existence would be when one of the terms (numerator or denominator) is positive and the other is negative. Then, even if $\left(\frac{p}{c}\right)^{\beta} \neq \frac{F}{F_e}$ and the curves do intersect, they would do so out of the positive quadrant, and thus not in our sphere of analysis. For this situation, there are two possibilities:

$$\frac{Y}{Y-k} < \left(\frac{p}{c}\right)^{\beta} < \frac{F}{F_e} \tag{4.23}$$

$$\frac{Y}{Y-k} > \left(\frac{p}{c}\right)^{\beta} > \frac{F}{F_e} \tag{4.24}$$

Given our assumptions, these two possibilities are plausible, since $F \geq F_e$, $p > c$, thus every term of the inequalities is larger than 1. Interestingly, all numerators refer to local consumption behaviour and all denominators to online shopping. Once it is established the condition for the two curves not to meet, it is worth exploring which behaviour will prevail throughout the landscape the landscape ($\Psi(x) > \Psi_e(x)$ or $\Psi(x) < \Psi_e(x)\ \forall x$). For that, we determine which behaviour will prevail in the centre ($x = 0$), since it can be applied for all x once the curves do not meet. With that definition, it will also be possible to determine the behaviour occupying the centre for the situations the curves do meet. The online-shopping behaviour will occupy the centre if $\Psi(0) < \Psi_e(0)$. For that to happen, after rearranging Equations 4.9 and 4.10, we reach the following condition:

$$\frac{Y}{Y-k} < \left(\frac{p}{c}\right)^{\beta} \tag{4.25}$$

Conversely, for the local consumption to occupy the centre, the following condition is necessary:

$$\frac{Y}{Y-k} > \left(\frac{p}{c}\right)^{\beta} \tag{4.26}$$

Intuitively, one can say that, if the importing fee (k) is too high, it offsets the financial benefit of a lower unitary price (c) upon importing, making the first term of the inequality larger than the second, and local consumption prevails (Equation 4.26). The same argument can be observed from the opposite perspective: if local prices (p) are too high compared to c, it can offset the financial benefits of not paying an importing fee, making the second term larger, and online-shopping prevails (Equation 4.25).

Once the curves do meet, it is important to observe which behaviour would occupy central areas, while the other occupies the peripheries. Knowing already the conditions for each behaviour to occupy the centre (Equations 4.25 and 4.26), it is important to add for each case a simultaneous condition with the opposite behaviour occupying the fringe. By first defining the fringes as the points f and f_e in the x-axis at which the bid-rent curves (Equations 4.9 and 4.10, respectively) intercept the agricultural rent constant (R_A), and thus the city ends, we have the following fringes defined:

$$f = \frac{Y}{F} - \frac{\bar{u}}{F}\left(\frac{p}{\beta}\right)^{\beta}\left(\frac{R_A}{\alpha}\right)^{\alpha} \tag{4.27}$$

$$f_e = \frac{Y-k}{F_e} - \frac{\bar{u}}{F_e}\left(\frac{c}{\beta}\right)^{\beta}\left(\frac{R_A}{\alpha}\right)^{\alpha} \tag{4.28}$$

With that, we can combine with whoever occupies the center to reach the next two land-use patterns. For online shoppers to occupy the centre with local consumers in the periphery, the two must be true: $\Psi_e(0) > \Psi(0)$ and $f > f_e$. By testing the second inequality and rearranging terms, we reach the following:

$$\frac{F}{F_e} < \frac{Y - \bar{u}\,(p/\beta)^{\beta}\,(R_A/\alpha)^{\alpha}}{Y - k - \bar{u}\,(c/\beta)^{\beta}\,(R_A/\alpha)^{\alpha}} \tag{4.29}$$

We can see that the right-hand side of the inequality is the previously mentioned disposable income ratio $\frac{Y}{Y-k}$ with both numerator and denominator reduced by constants, called respectively here A and B for simplicity. If the ratio between A and B is equal to the original disposable income ratio, the resulting ratio would be the same. If A/B is greater than the disposable income ratio, the resulting ratio will be smaller and vice-versa. As it happens, by dividing A/B we reach exactly the relative prices ratio $\left(\frac{p}{c}\right)^{\beta}$, which we know by Equation 4.25 is larger than the disposable income ratio in this situation. This means the resulting ratio of the right-hand side of Equation 4.29 will be even smaller than the original disposable income ratio. Since the trip-frequency ratio must be even smaller than that, we can safely say that:

$$\frac{F}{F_e} < \frac{Y}{Y-k} < \left(\frac{p}{c}\right)^{\beta} \tag{4.30}$$

This set of inequalities guarantees that the two curves will, necessarily, cross each other, and online-shoppers occupy the central area. If the ratio $\frac{F}{F_e}$ is too high, to the point where it surpasses the value of $\frac{Y}{Y-k}$, while still under the $\left(\frac{p}{c}\right)^{\beta}$ threshold, we break the fringe condition ($f > f_e$) and the curves do not meet. Moreover, if the ratio $\frac{Y}{Y-k}$ is too high, to the point where it surpasses the threshold value of $\left(\frac{p}{c}\right)^{\beta}$, we reach the condition of non-existence expressed by Equation 4.24. Conversely, for local consumers to occupy the centre with online shoppers in the periphery, we must have $\Psi_e(0) < \Psi(0)$ and $f < f_e$ simultaneously, leading to:

$$\frac{F}{F_e} > \frac{Y}{Y-k} > \left(\frac{p}{c}\right)^{\beta} \tag{4.31}$$

Similarly, but with an inverted argumentation, the value of $\frac{F}{F_e}$ acts as a topping threshold. If the ratio $\left(\frac{p}{c}\right)^{\beta}$ is too high, to the point where it surpasses the value of $\frac{Y}{Y-k}$, we reach the condition of nonexistence expressed by Equation 4.23. Moreover, if the ratio $\frac{Y}{Y-k}$ is too high, to the point where it surpasses the threshold, we break the fringe condition and the curves do not meet.

A final scenario, one of integration where the two bid-rent curves coincide for all x, is met when the two behaviours are indifferent towards their location in relation to the centre. This happens when the three ratios are the same:

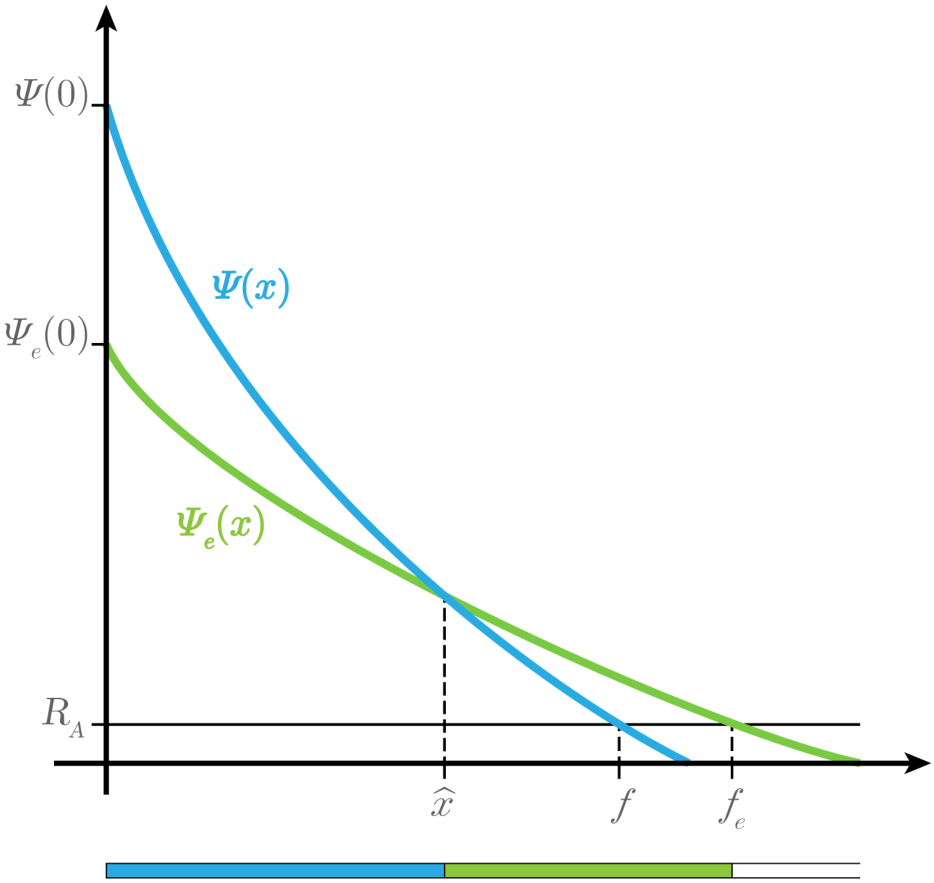
$$\frac{Y}{Y-k} = \left(\frac{p}{c}\right)^{\beta} = \frac{F}{F_e} \tag{4.32}$$

A summary of all possible land-use patterns, depending on consumer behaviour, according the parameter choice, is found in Table 4.4.1.

Several interpretations are possible with this definition of land-use patterns. Firstly, we must think of the parameter k as an overall correspondent to delivery costs. If we consider that delivery costs have been diminishing with time, it makes the disposable income ratio ($\frac{Y}{Y-k}$) smaller with time (closer to 1), making it more likely that local shoppers will occupy the central areas. If we take it to the extreme case, this would mean the most likely scenario would be with online-shoppers exclusively occupying the whole landscape.

These two likely scenarios are further reinforced when we consider the ratio between visits to the city centre ($\frac{F}{F_e}$). Departing from the innovation-diffusion theory, and based on studies that claim that higher-income professions are more likely to adopt teleworking (Wieland, 2023b), these higher-income professions are also more likely to be in the online-shopper category of our model. This would mean that

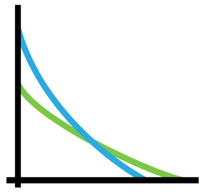
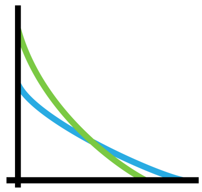
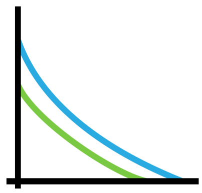
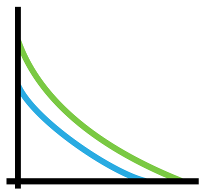
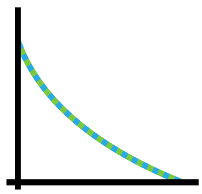
Figure 4.4.1: A visual representation of the different bid-rent curves for both behaviours, with their y-intercepts, intersects with agricultural rent (R_A) and between each other ($\hat{x}$). Lower bar represents dominating behaviour.



the visit to the city centre ratio would be even higher, further reinforcing the two previously mentioned scenarios of land-use distribution.

With all that in mind, we can highlight that the relative price ratio ($(\frac{p}{c})^\beta$) must be small enough so that local shoppers still exist at all in the landscape. This could be achieved by a higher competition in the city centre, that in an endogenous-price set-up would tend to drive prices p down. Conversely, an induced increase of k could also generate the same effect: this would be the case, for instance, for further taxation of deliveries, a discussion that might interest governments that would like, from a political economy perspective, to foster the local shops.

Table 4.4.1: Summary of land-use patterns and respective conditions of existence.

Pattern	Description	Conditions	Illustration
Local shoppers in the centre	$\Psi(0) > \Psi_e(0)$ $f < f_e$	$\frac{F}{F_e} > \frac{Y}{Y-k} > \left(\frac{p}{c}\right)^\beta$	
Online shoppers in the centre	$\Psi(0) < \Psi_e(0)$ $f > f_e$	$\frac{F}{F_e} < \frac{Y}{Y-k} < \left(\frac{p}{c}\right)^\beta$	
Strictly local shopping	$\Psi(x) > \Psi_e(x) \forall x$	$\frac{Y}{Y-k} > \left(\frac{p}{c}\right)^\beta > \frac{F}{F_e}$	
Strictly online shopping	$\Psi(x) < \Psi_e(x) \forall x$	$\frac{Y}{Y-k} < \left(\frac{p}{c}\right)^\beta < \frac{F}{F_e}$	
Integrated be- haviours	$\Psi(x) = \Psi_e(x) \forall x$	$\frac{Y}{Y-k} = \left(\frac{p}{c}\right)^\beta = \frac{F}{F_e}$	

4.4.2 Discontinuities in the urban fabric

Given the existence of a point $\hat{x}$ at which the two bid-rent curves cross, it is crucial to investigate the implications of the two different behaviours on urban fabric patterns. Firstly, the consumption of housing and, by direct consequence, the density curve of the landscape. Even if, by definition, the bid-rent at point $\hat{x}$ is equal for the two behaviours, that does not necessarily mean the housing consumption will be the same. In fact, it can be seen that the relations between relative prices play an important role on the outcome landscape, which is investigated in this section.

Given the curves S^* and S_e^* previously defined, we define the difference between them at $x = \hat{x}$, i.e., $S^*(\hat{x}) - S_e^*(\hat{x})$, as $\Delta S^*(\hat{x})$. If the difference is positive, it means that at the intersection point the locally consuming households consume also more housing than their importing counterparts, and vice-versa. We, thus, move towards investigating what are the conditions for that difference to be positive or negative.

$$\Delta S^*(\hat{x}) = \frac{\alpha\,(Y - t(\hat{x}))}{r(\hat{x})} - \frac{\alpha\,(Y - t_e(\hat{x}) - k)}{r(\hat{x})} \tag{4.33}$$

By substituting their respective bid-rents and transportation costs at $\hat{x}$, we reach:

$$\Delta S^*(\hat{x}) = \bar{u}^{(1/\alpha)} \left(\left((Y - F\hat{x}) \left(\frac{\beta}{p} \right) \right)^{-\frac{\beta}{\alpha}} - \left((Y - F_e\hat{x} - k) \left(\frac{\beta}{c} \right) \right)^{-\frac{\beta}{\alpha}} \right) \tag{4.34}$$

From there, we can substitute the value of $\hat{x}$ defined by Equation 4.22. After some rearrangement, we reach the following form, that allows for deriving conditions of positivity. For ease of readability, we substituted the term $\left(\frac{p}{c}\right)^{\beta}$ for A.

$$\Delta S^*(\hat{x}) = \bar{u}^{(1/\alpha)} \left(\frac{(AF_e - F)}{\beta\,(F_eY - FY + Fk)} \right)^{\frac{\beta}{\alpha}} \left(\left(\frac{p}{A} \right)^{\frac{\beta}{\alpha}} - c^{\frac{\beta}{\alpha}} \right) \tag{4.35}$$

Given the first term in the equation, $\bar{u}^{(1/\alpha)}$, is positive by assumption, the other two terms must be either both positive or both negative for the difference to be positive and local consumers will consume more housing at $\hat{x}$. Conversely, if either of the terms is positive while the other is negative, the difference will be negative and online shoppers will consume more housing at $\hat{x}$.

For the difference to be positive, we test first for the third term of Equation 4.35 being positive. We can see that, by rearranging and substituting $\left(\frac{p}{c}\right)^{\beta}$ back in place of A, that term will be positive when:

$$\frac{p}{c} > \left(\frac{p}{c}\right)^{\beta} \tag{4.36}$$

Since, by assumption, $\beta \in (0, 1)$ and $\frac{p}{c} > 1$, the inequality defined in Equation 4.36 will always be true and, thus, the third term in Equation 4.35 will always be positive. It remains to be seen the behaviour of the second term to define either the overall difference will be positive or negative.

For the second term to turn out negative, either the numerator or the denominator of the second term must be negative, while the other remains positive, or vice versa. For the numerator to be strictly positive, when substituting $\left(\frac{p}{c}\right)^{\beta}$ back in place of A, we reach the following condition:

$$\frac{F}{F_e} < \left(\frac{p}{c}\right)^{\beta} \tag{4.37}$$

Simultaneously, the condition for the denominator to be strictly negative, knowing that $\beta \in (0, 1)$, yields the following condition:

$$\frac{Y}{Y-k} < \frac{F}{F_e} \tag{4.38}$$

This second condition indicates that, according to Equations 4.27 and 4.28, $f_e > f$ and online-shoppers occupy the periphery. However, for the point $\hat{x}$ to exist in this situation, local shoppers must occupy the centre, requiring the condition set by Equation 4.26. The combination of conditions set by inequalities 4.37 and 4.38 contradicts that. This is also true for the reverse reasoning of said conditions, i.e., for the numerator to be strictly negative while the denominator is strictly positive. We can infer, then, that the difference in housing consumption set by Equation 4.35 will always be positive whenever a single point $\hat{x}$ does exist. In other words, the only discontinuity in the urban fabric possible to occur is by local shoppers consuming more housing than online shoppers at point $\hat{x}$, never the other way around. By extension, density will be higher for online shoppers than for local shoppers at that same point.

4.4.3 Profitability of central retail

One of the main discussions concerning the future of retail is what central retail spaces will be like with the expansion of online shopping or, more broadly, with the parallel diminishing of overall retail consumption as online-shopping expands (Wieland, 2023b). In this model, this can be seen as a reduced number of profitable firms in the city centre. Since we incorporate an observable competition between an external omnipresent good provider and local shops, we can check under which configurations one can affect the other. To do that, we test first what are the parameter conditions for central retail to be profitable at all. We then move into the dynamics of the optimum number of firms in the city centre: whether it is possible, given the increasing population, characteristic of this open model, for the optimum number of firms to decrease at some point. This would mean that a mere population increase would not suffice for higher consumption in the centre, depending more on how that new population would behave in terms of their shopping habits or indirect effects of that increase of population.

Given the assumptions and set-up configurations of our model, we can first infer the channel via which such phenomenon could occur: increasing rents. We depart from the hypothesis that, given a continuously increasing population, the diminishing achieved utility $\bar{u}$ at the fringe $\bar{x}$ will provoke rent increases in central retail locations up to a point where it offsets their potentially larger customer base from this increase in population. We investigate the possibilities for that to happen in two main scenarios. Firstly, supposing local shopping does happen in the model, it is assumed to be more likely that it will happen near the city centre, as shown before. This leads us to two scenarios: online shoppers in the centre and strictly local shopping. For each of these scenarios, we first show what parameter configuration is needed for central shops to be profitable, and then whether incoming residents can pressure rent in the centre up to a point the total number of shops starts going down.

To investigate the profitability of shops in the centre, we modify Equation 4.21 from a discrete aggregation of agents towards a continuous one, from the centre (0) to the fringe of local shoppers (f), considering both that $f < \hat{x}$ and $f < \bar{x}$, i.e., f is smaller than both the absolute fringe and a possible land-use frontier:

$$Q' = \int_0^f Z^* \tag{4.39}$$

With the assumption local shoppers will be near the centre, it is their bid-rent that is prevailing there (Equation 4.9), that shops need to outbid. Since transportation costs at the centre equal 0, the term $t(x)$ is zeroed, leading to the following rent at the centre:

$$r(0) = \alpha \left(\frac{Y}{\bar{u}}\right)^{1/\alpha} \left(\frac{\beta}{p}\right)^{\beta/\alpha} \tag{4.40}$$

By substituting both equations into 4.20, we reach the number of firms in the centre dependent on the utility reached at the fringe, $\bar{u}$. To test for the profitability, we must say that this number needs to be higher than 0, leading us to:

$$\frac{\beta\,(p-c)\left(Y^2 - \left(\bar{u}\left(\frac{R_A}{\alpha}\right)^{\alpha}\left(\frac{p}{\beta}\right)^{\beta}\right)^2\right)}{2Fp\left(\alpha\left(\frac{Y}{\bar{u}}\right)^{1/\alpha}\left(\frac{\beta}{p}\right)^{\beta/\alpha} + k\right)} > 0 \tag{4.41}$$

Since all the other terms of the equation are positive by assumption, it leaves us with the last term of the numerator to be also positive for the whole equation to be positive overall. After some rearrangement, we reach a non-sufficient condition for profitability for central retail in this situation:

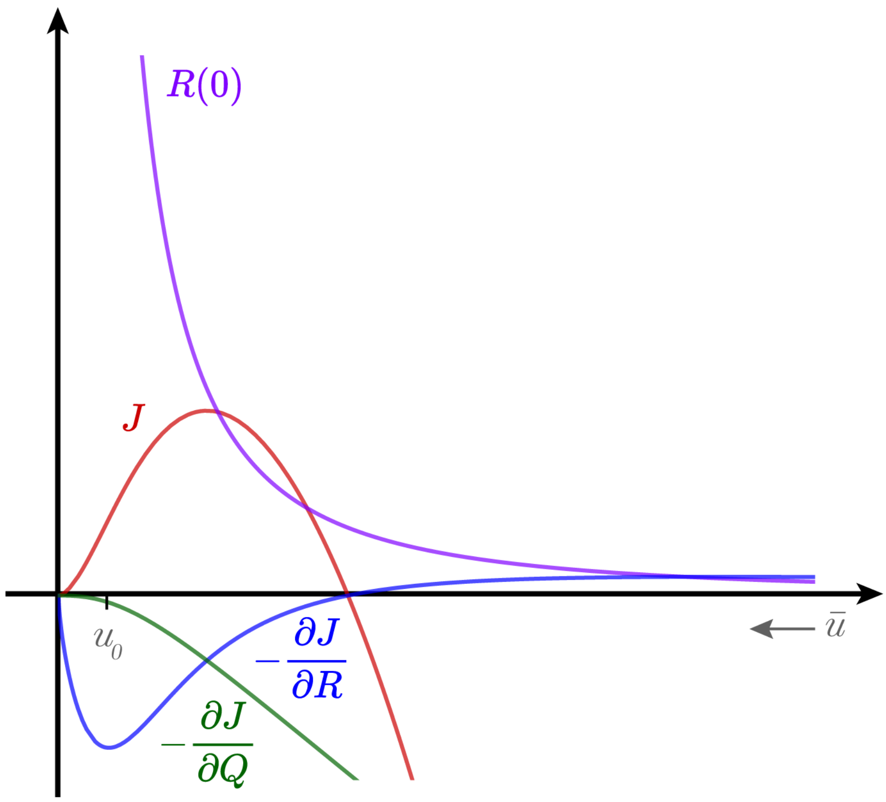
$$\bar{u} < Y\left(\frac{\alpha}{R_A}\right)^{\alpha}\left(\frac{\beta}{p}\right)^{\beta} \tag{4.42}$$

Departing from that, it is now vital do discuss whether profits in the centre, proxied in this model by the total number of firms present in the city centre, could diminish with time. Like previously mentioned, it could be the case that an added household to the fringe will drive rents higher than the added benefits to the consumption. This would mean in a situation where, for some choice of parameters, a marginal addition of firms is countered by a marginal reduction of firms with a marginal reduction in utility at the fringe $\bar{u}$:

$$\frac{\partial J}{\partial \bar{u}} = 0 \tag{4.43}$$

This means that the marginal increase in number of firms due to increasing quantity sold is completely compensated by the marginal decrease in number of firms due to an increase in rents. To illustrate this relation, we created a chart for the situation where strictly online shopping happens. We can see, illustratively in Figure 4.4.2, that the number of firms in the centre peaks at where the marginal rates equate. As the population in the open model increases, the utility at the fringe $\bar{u}$ decreases until it reaches the outer-world (u_0).

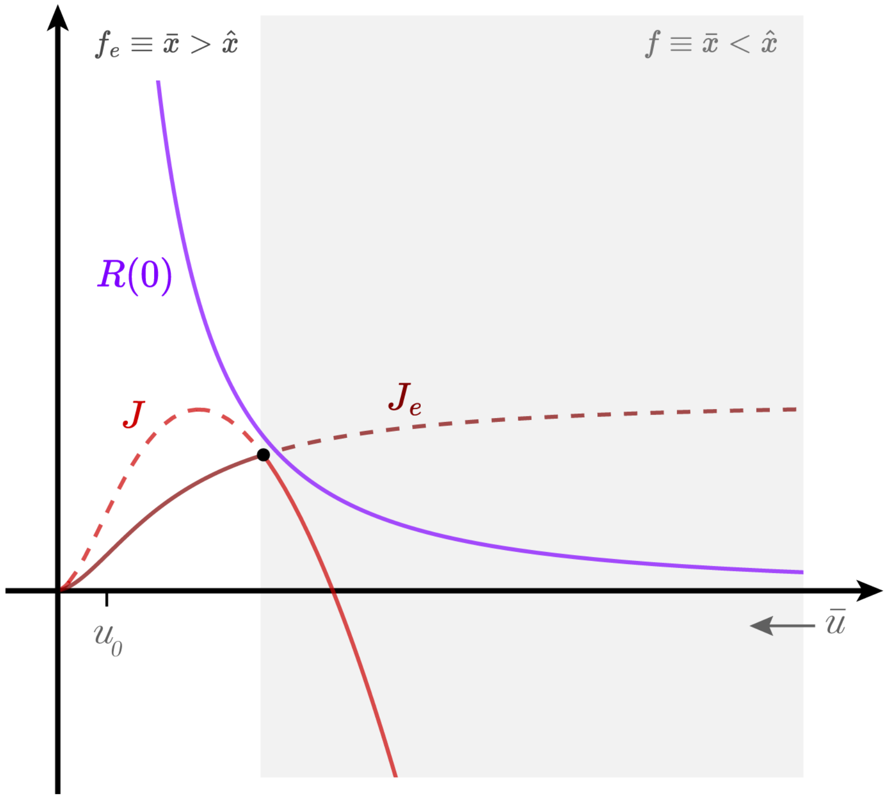
Figure 4.4.2: Optimum number of firms (J) in the centre as the population increases for a scenario of strictly local shopping.



Now turning to a situation where both online-shopping and local shopping happen in the model, we consider the more likely scenario in which local shopping locates at the centre. As the city expands (diminishing $\bar{u}$), the city is first occupied with only local shoppers, at which point the absolute fringe equals the local shoppers' fringe ($\bar{x} = f$), smaller than a theoretical intersection point $\hat{x}$. At some point, the absolute fringe equals the intersection point, and, after that, local consumption happens in the interval $[0, \hat{x}]$ and the absolute fringe equals the online-shoppers fringe ($\bar{x} = f_e$). At this point, we modify the total quantity consumed (Equation 4.39), and the total number of firms is now named J_e. All this is summarised in Figure 4.4.3.

$$Q'' = \int_0^{\hat{x}} Z* \tag{4.44}$$

Figure 4.4.3: Optimum number of firms (J) in the centre as the population increases for a scenario where, first strictly local shopping happens followed by online-shoppers in the periphery.



We can see from the illustration in Figure 4.4.3 that, for the parameter configuration where local-shoppers inhabit the centre, the curve J equals that of Figure 4.4.2. However, when the absolute fringe $\bar{x}$ equals the intersection point $\hat{x}$, the optimum number of firms in the centre is now governed by J_e, by quantity defined in Equation 4.44. We can immediately see that, as compared to the other situation, the number of firms will necessarily reach a lower peak, exactly at point where the intersection $\hat{x}$ equals the absolute fringe $\bar{x}$. This means that no online-shoppers will be inhabiting the landscape yet. With online-shoppers moving in, the optimum number of firms will necessarily be less than without them.

4.4.4 External shock: teleworking

In this section, we investigate whether an external shock that increases teleworking has the capacity to significantly alter the constituted built landscape. Departing from the learned collective lessons of a pandemic, at which work-from-home became the standard for a large parcel of workers from one day to the other, we can see this as a not unlikely shock scenario. An increase in teleworking diminishes the frequency of visits to the city centre, where employment is provided in this model, altering the composition of the transportation costs of households, and subsequently, their bid-rent curves. We test how strong these alterations need to be so that the landscape reaches other equilibria.

First we analyse if a constant C is reduced from both F and F_e, meaning the teleworking shock affects equally both behaviours and they will equally reduce their number of visits to the city-centre. Given that, by assumption, $F > F_e$, if both numerator and denominator are reduced by the same constant, the only possibility is that the resulting ratio $\frac{(F-C)}{(F_e-C)}$ will be greater than the original ratio. For two of the scenarios, according to table 4.4.1, this would not change anything. This is the case for when local shoppers are in the centre, or strictly online shopping dominates. In these two cases, the ratio is at its highest, and increasing it would not alter the composition of the landscape.

On the other hand, for the scenario where strictly local shopping happens, an increase in the $\frac{F}{F_e}$ ratio can lead it straight to a scenario where local shoppers are in the centre. Effectively, with the flattening of both bid-rent curves, the online-shoppers fringe (f_e) increases even more intensively, generating an urban sprawl of predominantly online-shoppers. For the scenario where online-shoppers are in the centre, an increase in the $\frac{F}{F_e}$ ratio could lead only to a scenario where strictly online-shoppers dominate the landscape, for the same reason described above. In other words, a shock in teleworking could only lead to a situation of (1) longer urban fringes, and (2) with more online-shoppers than before. Given that both phenomena affect the profitability of central retail, it can be said that this would have a direct impact on the number of firms in the city centre.

For a different reduction from F and F_e, meaning the teleworking shock affects differently both behaviours, it could be expected from literature that professions that allow for teleworking are, on average, with higher income and, thus, more prone to online-shopping. This would mean that the reduction from F_e, C_e would be higher than that of F, C. This would only further reinforce the land-use pattern changes already discussed in the previous paragraph, as this would mean the change in the ratio would be even higher.

4.5 Conclusion

The first discussion arising from the implementation of the model is that diminishing delivery costs tend to reduce the viability of central retail and significantly alter the behaviour of those inhabiting the city. Secondly, for the more likely scenarios in terms of parameter composition, we see that the presence of an online-shopping behaviour tends to increase the absolute fringe of a city. That reinforces the expectation already described in literature and hypothesised prior to the conducting of this work: that geographically sticky burdens, such as transportation and delivery costs, tend to agglomerate people and firms, whereas their diminishing tend to dispersion. This dispersion being driven mainly by the online-shopping behaviour indicates that it needs to be addressed as having public consequences: a more sprawled city tends to increase negative externalities for all who live in it.

Moreover, the model discussed how the viability of retail in the city centre is affected by online-shopping even in an indirect way: via rent increases. Of course, it is also something that can be intuitively thought, but observing the specifics at which this happens adds an extra layer to the discussion. Even if nothing changes for those who consume locally - they will still commute normally, and shop the same amount; shops can become unviable if incoming households tend to prefer to shop online. These households will put pressure on rent, by being further away from the city centre, and will shop online following specially the efficiency hypothesis. Effectively, online-shopping by itself as it is implemented in the model, reduces the viability of central retail.

Finally, it can be observed also how habits of teleworking impacts the land-use configuration of this theoretical city. By reducing the number of visits to the city centre, it affects the viability of central retail directly, but also indirectly: by altering the dominating shopping behaviour once households do not have to commute that often. In other words, by altering the land-use equilibrium of the model. It is also interesting to observe how differing the impact of teleworking on both behaviours generates the same outcome. The intuition for this is the following: those who most often telework tend to be higher-skilled, which also coincides with those who perform the most online shopping. This further reinforces the expected land-use changes from when both behaviours are equally impacted. It also brings this modelling approach closer in scope to the 'information-diffusion' hypothesis, although not predominant.

Overall, the model developed in this research complements the discussion of the impacts of online-shopping on the urban form. It incorporates the online-shopping behaviour and the possibility of teleworking into a monocentric city model. The differing bid-rent curves for different preferences govern the land-use distribution of this theoretical city and is sensitive specially to importing fees, relative prices, and frequencies of visits to the centre. As in any theoretical work, it is thought as a base for enhancing discussions by simplifying dynamics of a complex topic. Future work is encouraged to expand the dynamics into two-dimensions, allowing firms to leave the city centre, as it will introduce the layer of suburban shopping to the equation. Furthermore, parameter estimation and calibration in future work could also expand the understanding of applicability of the results discussed.

References

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press.

Althoff, L., Eckert, F., Ganapati, S., & Walsh, C. (2022). The Geography of Remote Work. *Regional Science and Urban Economics*, *93*.

Anderson, W. P., Chatterjee, L., & Lakshmanan, T. (2003). E-commerce, Transportation, and Economic Geography. *Growth and Change*, *34*(4), 415–432.

Araldi, A., & Fusco, G. (2019). Retail Fabric Assessment: Describing retail patterns within urban space. *Cities*, *85*, 51–62.

Arranz-López, A., Andreas, B., & Martin, L. (2023). Exploring the associations between E-shopping and the share of shopping trip frequency and travelled time over total daily travel demand. *Travel Behaviour and Society*, *31*, 202–208.

Barrero, J. M., Bloom, N., & Davis, S. J. (2021, July 21). Let Me Work From Home, or I Will Find Another Job.

Beckers, J., Birkin, M., Clarke, G., Hood, N., Newing, A., & Urquhart, R. (2022). Incorporating E-commerce into Retail Location Models. *Geographical Analysis*, *54*(2), 274–293.

Beckers, J., Weekx, S., Beutels, P., & Verhetsel, A. (2021). COVID-19 and retail: The catalyst for e-commerce in Belgium? *Journal of Retailing and Consumer Services*, *62*, 102645.

Brueckner, J. K., Kahn, M. E., & Lin, G. C. (2023). A New Spatial Hedonic Equilibrium in the Emerging Work-from-Home Economy? *American Economic Journal: Applied Economics*, *15*(2), 285–319.

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*, *46*(4), 563–585.

Chu, C.-P., Guo, W.-C., & Lai, F.-C. (2012). On the competition between an online bookstore and a physical bookstore. *NETNOMICS: Economic Research and Electronic Networking*, *13*(3), 141–154.

Colombo, S., & Hou, Z. (2021). Location-Price Equilibria when Traditional Retailers Compete Against an Online Retailer. *Review of Industrial Organization*, *59*(3), 483–502.

Farag, S., Schwanen, T., Dijst, M., & Faber, J. (2007). Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. *Transportation Research Part A: Policy and Practice*, *41*(2), 125–141.

Florida, R., Rodríguez-Pose, A., & Storper, M. (2021). Cities in a post-COVID world. *Urban Studies*, 00420980211018072.

Fujita, M., & Ogawa, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations, 36.

Gamber, W., Graham, J., & Yadav, A. (2023). Stuck at home: Housing demand during the COVID-19 pandemic. *Journal of Housing Economics*, *59*.

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, *1*(1), 27–50.

Glaeser, E. L., Luca, M., & Moszkowski, E. (2023). Gentrification and retail churn: Theory and evidence. *Regional Science and Urban Economics*, *100*, 103879.

Guo, W.-C., & Lai, F.-C. (2017). Prices, Locations and Welfare When an Online Retailer Competes with Heterogeneous Brick-and-Mortar Retailers. *The Journal of Industrial Economics*, *65*(2), 439–468.

Gupta, A., Mittal, V., Peeters, J., & Van Nieuwerburgh, S. (2021). Flattening the curve: Pandemic-Induced revaluation of urban real estate. *Journal of Financial Economics*.

Gupta, A., Martinez, C., & Van Nieuwerburgh, S. (2023, August). Converting Brown Offices to Green Apartments.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, *39*(153), 41–57.

Huff, D. L. (1966). A Programmed Solution for Approximating an Optimum Retail Location. *Land Economics*, *42*(3), 293.

Lai, F.-C. (2023). An Urban Configuration with Online Competition. *The B.E. Journal of Theoretical Economics*, *23*(2), 639–662.

Mills, P. E. S. (1972, June 1). *Studies in the Structure of the Urban Economy* (First Edition). John Hopkins Press.

Muth, R. M. (1969, January 1). *Cities and Housing: The Spacial Pattern of Urban Residential Land Use*. Univ of Chicago Pr.

Ogawa, H., & Fujita, M. (1980). EQUILIBRIUM LAND USE PATTERNS IN A NONMONOCENTRIC CITY. *Journal of Regional Science*, *20*(4), 455–475.

Orr, A. M., Stewart, J. L., Jackson, C., & White, J. T. (2023). Not quite the 'death of the high street' in UK city centres: Rising vacancy rates and the shift in property use richness and diversity. *Cities*, *133*, 104124.

Rao, F. (2019). Resilient Forms of Shopping Centers Amid the Rise of Online Retailing: Towards the Urban Experience. *Sustainability*, *11*(15), 3999.

Souza, T., Miyahira, R., Matheus, J., Nogueira, T., Maragoni-Santos, C., Barros, F., Costa Antunes, A., & Fai, A. (2022). Food services in times of uncertainty: Remodeling operations, changing trends, and looking into perspectives after the COVID-19 pandemic. *Trends in Food Science and Technology*, *120*, 301–307.

Teixeira, R., Duarte, A. L. d. C. M., Macau, F. R., & de Oliveira, F. M. (2022). Assessing the moderating effect of brick-and-mortar store on omnichannel retailing. *International Journal of Retail & Distribution Management*, *50*(10), 1259–1280.

Teulings, C., Ossokina, I., & Svitak, J. (2017). The Urban Economics of Retail. *The Urban Economics of Retail*.

Torrens, P. M. (2023). Agent models of customer journeys on retail high streets. *Journal of Economic Interaction and Coordination*, *18*(1), 87–128.

Ushchev, P., Sloev, I., & Thisse, J.-F. (2015). Do we go shopping downtown or in the 'burbs? *Journal of Urban Economics*, *85*, 1–15.

Vaughan, L. (2015). *Suburban Urbanities*. UCL Press.

von Thünen, J. (1863). *The Isolated State in Relation to Agriculture and Political Economy: Part III: Principles for the Determination of Rent, the Most Advantageous Rotation Period and the Value of Stands of Varying Age in Pinewoods*. Springer.

Wieland, T. (2023a). Pandemic Shopping Behavior: Did Voluntary Behavioral Changes during the COVID-19 Pandemic Increase the Competition between Online Retailers and Physical Retail Locations? *Papers in Applied Geography*, *9*(1), 70–88.

Wieland, T. (2023b). A Micro-Econometric Store Choice Model Incorporating Multi- and Omni-Channel Shopping: The Case of Furniture Retailing in Germany. *Geographical Analysis*, *55*(1), 3–30.

Young, M., Soza-Parra, J., & Circella, G. (2022). The increase in online shopping during COVID-19: Who is responsible, will it last, and what does it mean for cities? *Regional Science Policy & Practice*, *14*(S1), 162–178.

Chapter 5

Simulating retail location in the digital era: a spatially-explicit agent-based monocentric city model

Retail location competition is further complexified in the digital age, with the establishment of competitive online shopping alternatives. This study proposes a spatially-explicit agent-based model (ABM) to explore retail location dynamics in a two-dimensional monocentric city framework, integrating classical economic theories and recent challenges, such as digital commerce. We show that firms' location choices and economic viability are significantly influenced by the interplay between importing fees, income levels, and the opportunity cost of land. The model also bridges the principles of minimum differentiation, as proposed by Hotelling (1929), and maximum differentiation, by d'Aspremont et al. (1979), testing emergent patterns under varying transportation cost structures - linear and exponential. Findings reveal that both minimum and maximum differentiation can coexist, even in a pure exponential transportation cost implementation - where pure maximum differentiation would be expected otherwise. The study contributes methodologically to the field of location competition and urban economics by incorporating agent-based

simulations to capture the complexities of retail location decision in a dynamic urban environment, where a large number of agents (households and firms) interact.

5.1 Introduction

The retail landscape is undergoing a profound transformation driven by technological advancements, diminishing transportation costs, and shifting consumer behaviours. The proliferation of digital commerce has introduced new competitive dynamics, challenging traditional brick-and-mortar retailers and reshaping urban spatial structures (Beckers et al., 2021, 2022). Over the past decades, the diffusion of the automobile has induced a suburbanisation of population, due to reduced commuting costs. Given the role retail spaces have as places of vitality in urban centres, offering places of encounter and experiences in addition to the products, understanding the impact of these transformations on location patterns becomes crucial.

Spatial patterns of both firm and household location is known to be influenced by their behaviours. The common approach in urban economic models is to establish behaviour rules, given stylised facts, and analyse outcome patterns. Behaviours are constrained by costs, which yields a constrained optimisation problem for agents. The aggregate solving of these generate equilibria, which are then analysed in face of changing phenomena. However, traditional modelling approaches, often over a linear city, may fall short of capturing some key aspects of location competition in the retail sector. Firstly, the chaining of trips (commuting and shopping) together with varying transportation cost functions may generate non-circular, asymetric configurations of optimum location. Secondly, indirect interactions, when firms define prices depending on their competitors', may generate spatial externalities and path-dependencies that propagate patterns through time in a dynamic model. Finally, given the focus of this work on the importance of proximity, either to shops or to work, it is fundamental to extrapolate the one-dimensionality of traditional approaches to account for these.

This research aims to bridge traditional theoretical frameworks in urban economics (i.e., the monocentric city models, location-competition models) by developing a spatially explicit agent-based model (ABM) that explores retail location dynamics within a two-dimensional monocentric city. By integrating classical urban economics models with contemporary challenges posed by digital commerce, we seek to understand how varying transportation cost structures — both linear and exponential — affect the emergent spatial patterns of retail firms and households. We extend traditional location-competition models by considering an n-firm game and incorporating online shopping as a competitive force. This approach allows us also to investigate location patterns in a dynamic urban environment where numerous agents interact and make location and pricing decisions.

The primary objectives of this study are:

1. To examine where retail firms emerge and cluster in an open, agent-based, two-dimensional theoretical urban model.
2. To analyze how altering transportation costs and importing fees affects the equilibrium location of both households and firms.
3. To explore the location patterns that arise on the way to a price-location equilibrium in this setting.

To address these objectives, we employ agent-based simulations to capture the complexities, co-dependencies, and non-linearities inherent in retail location competition and urban spatial dynamics. This methodological approach enables us to model the interactions of numerous heterogeneous agents and to observe the emergent phenomena resulting from these interactions.

5.2 Background

From organisational studies to urban economics, different fields have tried to explain the different patterns of a firm's locational decisions. Seminally, Hotelling (1929) proposed a simple, linear model, in which two firms compete for a homogeneous market distribution by adapting their locations and prices. The outcome of these interactions would tend to attract both firms to the centre of the line, which was called the Principle of Minimum Differentiation (Hotelling, 1929), also known as Hotelling's law. Despite being thoroughly explored in further organisational studies and game theory literature, even being applied to non-spatial contexts - e.g., the political spectrum (Feldman et al., 2016), this outcome has also been questioned. d'Aspremont et al. (1979) claim that moving away from linear transportation costs breaks the stable equilibrium proposed by Hotelling (1929). In fact, given a certain set of conditions (e.g., quadratic transportation costs with distance), it could even lead to the opposite equilibrium: both firms having stable locations at the farthest point from one another (d'Aspremont et al., 1979), which was then called the Principle of Maximum Differentiation.

In a parallel trend of literature, Alonso (1964) proposed a simple linear model with microeconomic foundations that became known as the "monocentric city model". In its market equilibrium, households in the same city are equally satisfied by consuming less of the housing good as they locate closer to an exogenous city centre. This happens since they pay higher rent per unit area, which is offset by a lower total transportation cost. Lower housing consumption on the same unit of land generates a rising density gradient towards the unique city centre. This model has been further expanded onto what has become known as the Alonso-Müth-Mills (AMM) model.

There have been conciliations between the monocentric city literature and location competition models. Takahashi (2014) explicitly draws from both trends a model joining the same trade-offs of Alonso with retail competition over a line. Ushchev et al. (2015) explores location and price competition between central and suburban retailers. Other approaches over one dimension include Guo and Lai (2015), Guo and Lai (2020), Gupta et al. (2004), Kohsaka (1986), either at the linear city, or a circular city. Hotelling's linear city has also been expanded in two dimensions (Larralde et al., 2006) and with more than two firms (Fain, 2023). Differentiation in two dimensions, although not explicitly spatial ones, was also conducted by Irmen and Thisse (1998), highlighting different strengths for the differentiation axes matter.

Further specifications to either production costs (Dragone & Lambertini, 2020; Heywood et al., 2022), or transportation costs (Chen & Song, 2021) highlight the effects these can have on equilibrium locations. Anderson and Neven (1991) claims a Cournot-type oligopoly tends to lead to spatial agglomeration, whereas Netz and Taylor (2002) suggest market competition induces spatial differentiation. Another aspect commonly compared in literature is the actual order the game takes place, either location-then-price (Guo & Lai, 2013), or location-then-quantity (Hamilton et al., 1989). Most studies analysed here, however, restrict the analysis to a 2-firm game (Chen & Song, 2021; Dragone & Lambertini, 2020; Fleckinger & Lafay, 2010; Heywood et al., 2022), whereas for an n-firm game a key aspect is usually highlighted, such as how income differences influence equilibrium (Fain, 2023). Given the complexity of solving location and price equilibria for n-firms, computer simulation methods have also been found to be applied in literature (Gao et al., 2021). Beckers et al. (2021) emphasise the challenges faced by local retailers with a surge in e-commerce, incorporating e-commerce into traditional retail location models in a subsequent work (Beckers et al., 2022).

This research proposes to bridge Alonso's monocentric city model and retail location competition models by proposing a two-dimensional monocentric city in which competing retail firms and consuming households interact. We revisit the discussion between minimum vs. maximum differentiation principle, proposed initially by, respectively, Hotelling (1929) and d'Aspremont et al. (1979), by testing the differences between linear and exponential transportation costs. We extend traditional monocentric city models from a linear implementation to a spatially explicit, two-dimensional one. We also extend traditional implementations of location-competition from a limited number of firms to n-firms, endogenous to each setting. Finally, we introduce the relevant topic of online-shopping as an outside option for households in the model. The main objective of this work is to understand and explore the location patterns achieved by retail firms in equilibrium, sensitive to changes in transportation costs and online-shopping costs, as well as which particular dynamics lead to these equilibria. To account for non-linearities and path-dependencies, we choose to expand beyond an analytical and geometrical approach, towards modelling a spatially-explicit agent-based simulation, in which we test for different parameter configurations. This is done to address three main research questions:

1. Where do retail firms emerge and cluster in an open, agent-based, two-dimensional theoretical urban model?
2. How does altering transportation costs and importing fees affect the equilibrium location of both households and firms?
3. Which location patterns can be found on the way to a price-location equilibrium in this setting?

5.3 The model

We simplify the model defined in Chapter 4 to better understand the location decision and pricing mechanisms on a two-dimensional landscape. This also allows us to investigate different transportation functions, deemed essential for our research questions. In this section, we first start by defining household behaviour, moving on to firm behaviour, and specifications of the transportation functions, before following into equilibrium and comparative statics.

5.3.1 Household behaviour

In Chapter 4, households maximise utility, defined as a Cobb-Douglas function of housing consumption (S) and the homogeneous good (Z). For this work, we normalise housing consumption (S) to unity. It is a significant simplification, we justify it because the previous work already did not consider the main consequence of differentiated housing consumption to the urban landscape: density curves. By assuming for each point in space the density would be constant, to claim a household was consuming less than the unit did not make sense, since they could in practise expand their household to cover the whole lot they would be occupying, given no other restrictions are in place. Then, for this work, we consider $S = 1$, which yields, after a monotonic transformation, utility as equivalent to the consumption of homogeneous good Z.

$$U \propto Z \tag{5.1}$$

The utility maximisation problem is subject to a budget constraint, that concerns two types of household decisions: having the homogenous good delivered from the outside world, or buying it locally. If they buy locally, they incur an extra transportation cost on top of commuting. If they choose to import the good, they incur an extra delivery fee, k. This way, the decision to import serves as a baseline consumption for the household decision. As discussed in Magalhães & Caruso (2024), these can have a direct impact on the land-use patterns, which we will see later. The budget constraints, respectively for the local consumer and the online consumer, are, for each location x:

$$Y - t(x) = r(x) + Zp \tag{5.2}$$

$$Y - t_e(x) = r(x) + k + Z \tag{5.3}$$

with the importing product Z as the *numéraire*, $r(x)$ the rent paid at each location x. Since utility is simply the consumption of good Z, the utility-maximising Marshallian demand is the rearrangement of the budget constraints. Respectively for local and online consumers:

$$Z^* = \frac{Y - t(x) - r(x)}{p} \tag{5.4}$$

$$Z_e^* = Y - t_e(x) - r(x) - k \tag{5.5}$$

Given our simplified utility function, we can safely state the indirect utility of the two household decisions, respectively as $v(x) = Z^*$ and $v_e(x) = Z_e^*$. As a short-run equilibrium condition, all households have the same level of utility $\bar{u}$ equal to the highest achievable utility at the city fringe $\bar{x}$:

$$\bar{u} = \max\left(v(\bar{x}), v_e(\bar{x})\right) \tag{5.6}$$

This short-run equilibrium allows us to derive the respective bid-rents for both behaviours. We follow the bid-rent theory to determine the land-use. Whichever behaviour can pay a higher rent for the same utility for a given location x predominates. An initial intuition could point to the dominant behaviour for a location x being the one yielding the highest utility. However, bid-rent theory states that an absentee landowner would be able to charge higher for rent in that location for the utility premium a certain behaviour can yield. Thus, what dominates is the capacity to pay for that location given the fixed utility. In other words, a potential higher utility for a given location is compensated by a higher rent charged by the landlord, as other households would be willing to pay higher for that piece of land. This leads us to the two following bid-rent curves:

$$\Psi(x) = Y - t(x) - \bar{u}p \tag{5.7}$$

$$\Psi_e(x) = Y - t_e(x) - \bar{u} - k \tag{5.8}$$

Finally, the generalised residential rent function for any location x will be the highest between these two and an exogenously defined opportunity cost of land (called here the agricultural rent R_A):

$$r(x) = \max\left(\Psi(x), \Psi_e(x), R_A\right) \tag{5.9}$$

5.3.2 Transportation costs

The transportation cost specification differs from Chapter 4 by introducing exponential transportation costs with distance, in line with the utility framework literature (de Vries et al., 2004). The generalised transportation costs between two locations i and j is defined as follows:

$$T_{ij} = ce^{\gamma d_{ij}} \tag{5.10}$$

with c as an initial cost for travel, γ as the exponential parameter, and d_{ij} as the total traveled distance. We choose this specification since it allows us to explore both different spatial frictions (rate of increase of marginal transport costs with distance, parameter γ) and different fixed costs (transport costs for $d_{ij} = 0$, or parameter c).

All households commute to an exogenously defined Central Business District (CBD). Households will buy at a local shop (instead of importing) provided the extra transportation cost they incur with a shopping trip, summed with extra cost per product (given the firm's mark-up) will be smaller than the importing fee k (Equation 5.11).

$$t(x) - t_e(x) + \bar{u}\,(p - 1) \leq k \tag{5.11}$$

5.3.3 Firm behaviour

Shops in this model are trading firms, importing the uniform good by paying a unitary cost (*numéraire*) per product, and the importing fee k. They resell the product to households, given that their price p is sufficiently competitive for households' transportation costs (see Equation 5.11). This is possible due to economies of scale, as firms pay the same importing fee as the households, but serve a larger market, diminishing average costs. The quantity Q sold by each firm j will be:

$$Q_j = \sum_{i}^{I} Z^* \quad \forall \quad i \ni t_{ij}(x) \leq t_{ij'}(x) \tag{5.12}$$

where j' represents any other firm in the model. In other words, a household's consumption will add to a firm's quantity if the household's transportation cost to that firm is smaller than to any other firm.

A firm j's revenues and costs are defined as:

$$R(x) = Q(x)p \tag{5.13}$$

$$C(x) = Q(x) + k \tag{5.14}$$

The ability of a firm to install itself at a given location x depends on its bid-rent profile for any location x being at least higher than the highest household's. A representative firm's bid-rent Φ is defined as follows:

$$\Phi(x) = R(x) - C(x) \tag{5.15}$$

However, the effective rent paid by any firm at location x will be that of the second-highest bidder, either the households' or any other firm's. Intuitively, the absentee landowners capture the extra profit a certain location yields in the form of rent. This gives us the short-run equilibrium for firms, in which all firms have the exact same profit equal to that of the least-profitable firm ($\underline{\Pi}$). This defines the short-run profit for any location x as:

$$\Pi(x) = \Phi(x) - \max\left(r(x), \Phi(x, \underline{\Pi})\right) \tag{5.16}$$

Since firms are profit-maximisers, they will choose their price p^* such that their quantity sold is the optimum Q^* and their profit $\Pi^*(x)$ the maximum for that location. This happens when their marginal revenue equals their marginal cost, summed with the partial derivative of their rent with respect to quantity. This is defined in Equation 5.17. The optimum quantity $Q*$ is reached by solving the equation for Q, while the optimum price $p*$ by solving an inverse demand function of price as a function of quantity ($p(Q)$).

$$\frac{\partial R(x)}{\partial Q} = \frac{\partial C(x)}{\partial Q} + \frac{\partial \max\left(r(x), \Phi(x, \underline{\Pi})\right)}{\partial Q} \tag{5.17}$$

5.3.4 Equilibrium conditions

Some of the equilibrium conditions have already been described previously, but are summarised in this subsection.

1. **Short-run utility equilibrium:** The consumption level, which in this implementation represents utility itself, at any given time, must be the same for every household.

$$Z_e^i = Z^i = \bar{u} \quad \forall \quad i \tag{5.18}$$

2. **Short-run profit equilibrium:** The profit level of all firms, at any given time, must be the same for every firm.

$$\Pi_j = \underline{\Pi} \quad \forall \quad j \tag{5.19}$$

3. **Land-market equilibrium:** The effective rent paid at any location x must be the upper envelope of every bid-rent in the model, given the fixed utility and profit conditions.

$$r(x) = \max\left(\Psi(x, \bar{u}), \Psi_e(x, \bar{u}), \Phi(x, \underline{\Pi}), R_A\right) \tag{5.20}$$

4. **Consumption equilibrium:** The sum of the amount locally consumed by every household must equal the total quantity sold by every firm.

$$\sum_j Q_j = \sum_i Z^i \tag{5.21}$$

5. **Long-run equilibria:** In-migration stops when household utility reaches the outer-world utility, firm birth stops when firm profits reach 0.

$$\bar{u} \to u_0 \tag{5.22}$$

$$\underline{\Pi} \to 0 \tag{5.23}$$

5.4 Analytical & geometrical implications

We begin this section by simplifying initial assumptions, to be able to understand the effect of varying transportation costs, and exogenous variables such as income (Y), agricultural rent (R_A), importing fee (k), fixed transportation costs (c) and marginal transportation costs (γ) on the patterns of quantity ($Q*$) and price ($p*$) for each individual firm. We first linearise the model, by understanding over a linear city, with the CBD in the origin, what will be the patterns of the first and second firm, and how any two firms interact in relation to the CBD. Then, we highlight the dynamic dimension of the model, and how a gradual entrance of firms and households provide a specific set of tools to understand the different equilibria. Finally, we highlight the two-dimensionality aspect of the model, and how different positions in relation to the CBD (orthogonal or parallel) affect shopping behaviour and firm location strategies. Since all these interactions and possibilities increase greatly the complexity of the model, we justify the use of an agent-based spatially-explicit simulation to account for both the dynamics and the geometrical complexities of firms and households interacting in space.

5.4.1 Linear simplification

The first simplification is by linearising not only the landscape (towards a linear city), but also the transportation costs, to understand which behaviour (buying locally or importing) dominates. Borrowing the setting described in Chapter 4, we assume a higher marginal transportation cost for those who buy locally, hinting at a higher frequency of visit to the CBD than those who buy online. We also only consider scenarios where local buyers do exist.

First firm

We assume that the first firm will emerge very close to the CBD ($x \rightarrow 0$), as an initial configuration with only online-buyers will necessarily agglomerate symmetrically around the CBD. We adapt, therefore, the respective transportation costs to:

$$t(x) = 2\gamma_1 x \tag{5.24}$$

$$t_e(x) = 2\gamma_2 x \tag{5.25}$$

with $\gamma_2 < \gamma_1$ by assumption, and x being the distance from any point in the line to the CBD. The point where the two bid-rents intersect, $\hat{x}$, defines the point of indifference of a household towards buying from a shop in the centre or importing the good. The city fringe, $\bar{x}$, is defined as the highest between the two intersects between the residential bid-rents $(\Psi_e(x), \Psi(x))$ and the agricultural rent R_A. As defined in Equation 5.6 and as the short-run utility equilibrium, all households have the same level of consumption as that of the fringe. The intersect between the two bid-rents is, then, defined as:

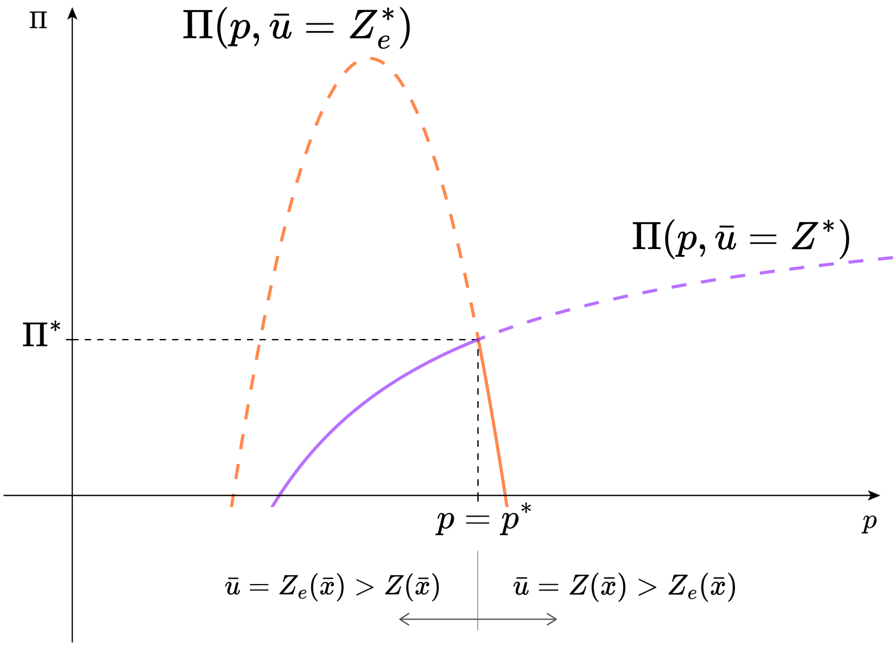
$$\hat{x} = \frac{\bar{u}\,(p-1) - k}{2\,(\gamma_2 - \gamma_1)} \tag{5.26}$$

which, for being positive, is in line with the indifference condition set by Equation 5.11.

It remains to be defined which is the dominant behaviour at the fringe, which will give us the equalised level of consumption $\bar{u}$. There can be two possibilities here: 1) where $\hat{x}$ is beyond both fringes ($\hat{x} > f > f_e$), when local-buying dominates the landscape and no online-shopping happens; 2) where $\hat{x}$ is smaller than both fringes ($\hat{x} < f < f_e$), when both consumptions happen, but local buying is closer to the CBD. The other possibilities discussed in Chapter 4 do not happen here by the assumptions laid before. In the first scenario, the fringe is composed by local-buyers, with $\bar{u} = Z(\bar{x})$; while in the second scenario, the fringe is composed of online-buyers, where $\bar{u} = Z_e(\bar{x})$.

For the first scenario, $\bar{u} = (Y - 2\bar{x}\gamma_1 - R_A)\,/p$ and the total quantity sold by the first firm is $Q_1 = \bar{x}\,(Y - 2\bar{x}\gamma_1 - R_A)\,/p$. Since its revenue is defined by $R_1 = Q_1 p$, price cancels out, leaving the revenue indifferent to price changes. The rent paid in the center is equal to that of the second-highest bidder, in this case local buyers themselves, equaling $\Psi(0) = Y - \bar{u}p$, which in this case simplifies to $\Psi(0) = 2\bar{x}\gamma_1 + R_A$, therefore rent is also indifferent to price changes. Its cost, however is equal to $C_1 = (Y - 2\bar{x}\gamma_1 - R_A)\,/p + k$, meaning an ever-increasing price would necessarily mean an ever-decreasing cost, leading its profit-maximising price to infinity. However, an ever-increasing price diminishes $\bar{u}$ to the point where it becomes lower than the

Figure 5.4.1: Profit curves as function of prices for both situations at the urban fringe $\bar{x}$.



online-consumption alternative ($Z_e(\bar{x})$) and the second scenario starts to dominate.

In the second scenario, $\bar{u} = (Y - 2\bar{x}\gamma_2 - R_A - k)$ and the quantity sold by the first firm is $Q_1 = \hat{x}\,(Y - 2\bar{x}\gamma_2 - R_A - k)$. Since $\hat{x}$ diminishes with price increases, with $\bar{u}$ not altering, revenue has forces acting in both directions. Rents tend to diminish with price increases, making it unclear which force will dominate. After solving for $\frac{\partial \Pi}{\partial p} = 0$, we reach a profit-maximising price so low that would switch again the dominating scenario in the fringe $\bar{x}$. Therefore, it is safe to say that the first firm's profit-maximising price will be the one that solves for an equal consumption level at the periphery ($Z(\bar{x}) = Z_e(\bar{x})$. In other words, a firm chooses their profit-maximising price low enough for the household at the fringe to buy locally, but not lower as it would start diminishing their profit by simply not attracting any new customers. We define it in Equation 5.27 and illustrate it in Figure 5.4.1.

$$p_1^* = \frac{Y - 2\bar{x}\gamma_1 - R_A}{Y - 2\bar{x}\gamma_2 - R_A - k} \tag{5.27}$$

Second firm

We test the entrance of a second firm in the line, located necessarily between the CBD (where the first firm is) and the fringe $\bar{x}$. The firm enters in the scenario described previously, with the existing firm in the centre practising the optimum price p_1^*. We follow the same logical that solved for the optimum price for the first firm, by adopting a consumption in the periphery ($\bar{x}$) that equalises that of the entering firm to that of online-shoppers ($Z_2(\bar{x}) = Z_e(\bar{x})$). This defines, first, the constant consumption $\bar{u}$, that is then used to calculate the optimum quantities and prices of the second firm.

The firm enters at a location $d \in [0, \bar{x}]$, and the transportation cost of any local consumer is defined as the sum of a commuting trip to the CBD, a shopping from there to the second shop, and a trip back home. Thus, their transportation cost is defined as follows:

$$t_2(x, d) = \gamma_1 (x + d + |x - d|) \tag{5.28}$$

This gives us the residential bid-rent for purchasing in the second firm, which when equal to that of the first firm yields the two intersection points, between which consumption to the second firm prevails. By accounting for the new constant consumption $\bar{u}$, and the price already practised by the first firm, these intersection points simplify to $\hat{x_1} = d - |\bar{x} - d|$ and $\hat{x_2} = d + |\bar{x} - d|$. From there, we define the firm's quantity and consequent profit $\Pi(d)$ and maximise it with respect to price p ($\frac{\partial \Pi}{\partial p} = 0$). We notice a similar behaviour to that of the first firm, in which the optimum price of the second firm is the one just enough to 1) make the consumer at the city's fringe $\bar{x}$ indifferent to either buying online, locally from the first firm, or from the second firm; and 2) the consumer at the centre ($x \to 0$) also indifferent between the first and the second firm, with equal bid-rents. This gives us the optimum price, according to distance d between firm and centre:

$$p_2^*(d) \leq \frac{Y - \bar{x}\gamma_1 - R_A - d\gamma_1 - \gamma_1|\bar{x} - d|}{Y - 2\bar{x}\gamma_2 - R_A - k} \tag{5.29}$$

We can see from comparing the optimum prices of the first and the second firm is that, for $d = 0$, they are the same. This means the second firm should practise a marginally lower price than the existing firm to maximise its profits, by monopolising the landscape for itself. This also means its profit-maximising position d is the minimum differentiation from the existing firm ($d \to 0$). By moving away from the city-centre, households in-between d and the CBD would necessarily want to purchase from the central firm, as this leg of their trip will happen anyway, as a commuting trip. For households between d and the city fringe $\bar{x}$, they are actually indifferent if the prices are the same, the reason why the second firm will practise a marginally lower price.

5.4.2 From static to dynamic

The analysis so far has been conducted in a simplified, linearised, static equilibrium, where firms and locations are based on the immediate configuration of the market and distribution of households. However, this framework does not capture the iterative nature of firm birth and exits, as suggested by the dynamic characterised for the second firm. It suggests that this implementation is, in fact, dynamic, with price-setting and firm locations evolving over time.

We can safely say that the previously discussed short-run equilibrium for both firm profits and household utilities already hinted at the dynamic characteristic of this model. More specifically, the profit achieved by the last-entering firm could be lower than the one already being practised by existing firms, setting a new baseline profit level ($\underline{\Pi}$) which triggers the readjustment of effective rents paid by firms. According to the bid-rent framework, a firm's higher profitability will be captured by the landowners in the form of rent, otherwise the entering firm would prefer to locate itself in its place. The same can be stated for households' short-run utility equilibrium, with the potentially higher utility for more accessible locations being captured by landowners in the form of rent, levelling the equilibrium utility ($\bar{u}$) by that of the city fringe (the furthest from the CBD). These conditions create a particular dynamic in which the reached profits and utilities of time-step s will be dependent on the ones achieved at $s-1$, thus, a dynamic model.

More specifically, the case for the second firm described in the previous subsection shows how their profit-maximising location will be marginally away from an initial central firm, while their price slightly lower. This strategy leads to a new monopoly arising every time a new firm enters, since it effectively steals the customers of the previous one. However, as time progresses, this dynamic is expected to repeat indefinitely, with each new firm marginally differentiating its location towards the side with more customers. As the city is continuously growing until reaching long-run equilibrium, this generates another dynamic phenomenon worth observing: in a discredited spatial framework like this one, firms act as if they are "moving" through the landscape, continuously substituting one another.

This “moving” phenomenon is expected to drive a continuous evolution of both firm and household locations. Over time, the cumulative effect of these adjustments could make it viable for a new firm to enter again the market near the CBD, now emptied by the described dynamic. For instance, as firms establish progressively away from the CBD, a new central point may again become the most profitable location for a new firm birth. This cyclical pattern, together with the corresponding optimum pricing strategies, suggests we are facing the dynamics of a complex system, for which agent-based simulations allow for greater insights to be drawn in the long run.

To formally capture these dynamic behaviours, we propose the introduction of a dynamic model, in which the market evolves over discrete time steps. At each time step s, a new household enters the landscape if there is a location x yielding an utility at least as much as an external utility u_0, either by consuming from any local shop or importing the good. The household enters paying the agricultural rent R_A and defines the fixed utility $\bar{u}$ that will trigger the readjustment of rents of all existing households, since any extra utility will be captured in rents. Therefore, all rents in time step s are defined by the fixed utility at $s-1$. Then, a firm will enter the landscape if there is any location yielding a positive profit. It will adopt the optimum price for that location, and households will switch to consume from it if their immediate utility will be higher. The profit by this last-entering firm ($\underline{\Pi}$) will trigger rent readjustments for all existing firms, since any extra profit will be captured in the form of rents. Therefore, all rents for firms in time step s are defined by the fixed profit at $s-1$. Finally, all rents and consumption decisions are updated according to the new location equilibrium, and any firms whose profit go negative will go bankrupt. Then a new time-step begins and this repeats until household utilities reach the external u_0 and profits reach 0.

5.4.3 From one to two dimensions

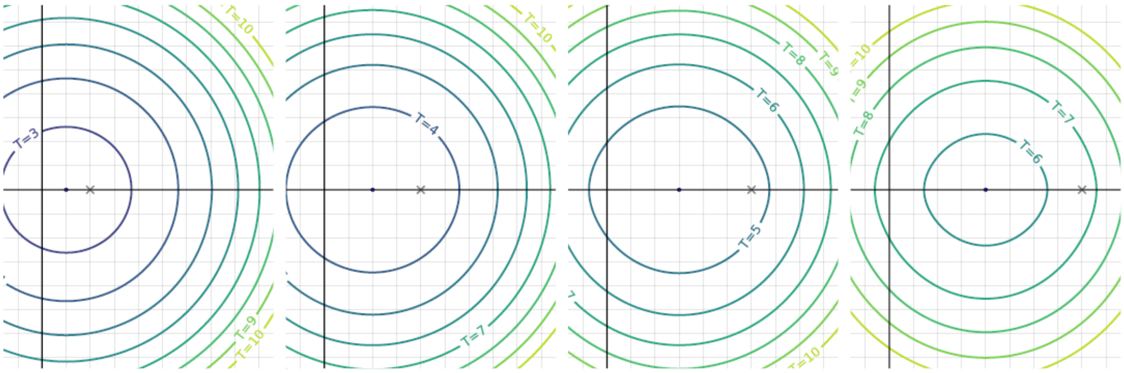
The analytical implications described over the city line, with linear transportation costs and interactions constrained to a single axis, provides valuables insights but does not capture the full extent of the expected dynamics. Firstly, the linear transportation costs fall short of capturing the different possibilities of location patterns trip-chaining is supposed to allow, as originally thought and described in the model definition. The reintroduction of exponential costs is expected to generate location equilibrium other than the minimum-differentiation principle, as highlighted by d'Aspremont et al. (1979) for non-linear transportation costs. Second, the linear setting explains the behaviour of competing firms orthogonal to the CBD, whereas parallel phenomena (i.e., when two firms are equally distant from the CBD) remain unaddressed. This is especially interesting in combination with the exponential transportation costs, as we expect to highlight a region of viability of firms, rather than a single point.

Firstly, we highlight that the possibility of trip-chaining over a line does not interfere with the consumption decision of the household located farther away than the firm. Since, in a line with a household in distance x from the CBD, and a firm in $d < x$, the total distant travelled is necessarily $2x$, for the household the firm's distance d from the CBD does not really make a difference. This is only the case because when $d + x + (x - d) = 2x$, which is the case for the line, their linear transportation costs combined $t(d) + t(x) + t(x - d) = t(2x)$ necessarily also holds. On top of that, for any firm location out of the line, all else equal, the option is also not considered as the transportation costs will necessarily be higher than to a shop on the line. This changes by reintroducing exponential transportation costs.

With exponential transportation costs, the costs of two smaller trips can be less than the cost of a single trip of both distances combined ($t(x) + t(d) + t(x - d) < t(2x)$, even when $x + d + (x - d) = 2x$). This can be the case even when the sum of the two smaller distances is greater than the longer distance: the total transportation cost can be smaller for combining two smaller trips than for the longer trip. This allows for a triangulation of trips, with shops located outside the line being less costly to access than ones on the line, closer to the household or to the CBD. We analyse the geometrical implications of that by looking at the household's side.

We pick a household located at distance x from the CBD. The transportation cost for any distance d is defined as in the model definition: $t(d) = ce^{\gamma d}$, with c acting as an initial cost, and γ as the marginal cost. Any household's total transportation costs will be the sum of a trip to the CBD $t(x)$, a trip from the CBD to the shop $t(d_1)$, and a final trip back home $t(d_2)$. We bring these to the Cartesian coordinate system, with CBD located at $(0, 0)$, household i located at $(x_i, 0)$ and firm j located at (x_j, y_j).

Figure 5.4.2: Contour lines for transportation costs given the location of a household in $x_i = 2$, $x_i = 4$, $x_i = 6$, $x_i = 8$, with minimum transportation costs highlighted at $x_i/2$.



This defines our distances $x = \sqrt{x_i^2 + 0^2}$, $d_1 = \sqrt{x_j^2 + y_j^2}$, $d_2 = \sqrt{(x_j - x_i)^2 + y_j^2}$. We fix different values for $\bar{T} = t(d_1) + t(d_2) + t(x_i)$ and solve for positions d_2 that reach the same transportation costs. We highlight these in the form of a contour plot, in which we vary x_i from 2 to 8 (Figure 5.4.2).

We highlight the quasi-elliptical shape of the contour lines, with the isoquants emanating from a centre point of minimum transportation cost exactly at $(x_i/2, 0)$, and foci at the CBD and the household location. For any point in the Cartesian plan, a firm at $(x_i/2, 0)$ would yield the lowest total transportation cost for a household at $(x_i, 0)$. The contour lines indicate possible firm locations with equal total transportation costs for the household in question. The quasi-elliptical shape of the isoquants differ from perfect ellipses precisely due to the exponential component.

5.5 Simulation

This section details further the implementation of the agent-based simulation of the model described so far. We implement it to test for unexpected patterns or complex outcomes that the theoretical, analytical and geometric approach fall short of addressing. These include the dynamic subsequent substitution of firms seeking minimum differentiation, which we called the "moving" dynamic; the subsequent expected maximum differentiation, of a firm again appearing farther from existing ones; the isoquants of exponential travel costs, generating regions of viability rather than single points; the co-dependence of optimum prices of parallel firms on each other. It is important to highlight that this is a non-exhaustive list, and this research remained open to identifying unexpected phenomena as well.

The simulation helps answer the research question of under which conditions, methodological or parameter-wise, do retail firms suburbanise and become more accessible to peripheral households. Departing from the setting described in Chapter 4 in which firms are only possible in the CBD, we have highlighted, analytically, that the lowest transportation cost point for households would only be out-of-the-centre in a situation of exponential transportation costs. This is further confirmed by the benchmark model, defined in the following subsections. We choose to define a benchmark model, with a choice of parameter values that expects, for small variations, significant changes in observable patterns.

The implementation of the simulation is done in *python*, with each agent defined as a separate instance of a pre-defined class with certain attributes. The environment consists of a series of *numpy* two-dimensional arrays, of size (X, Y), from which agents extract and impose attributes. These attributes are: effective rent, each one of the bid-rent curves, number of agents in a cell (limited to 1 in this setting), potential utility, potential profit, profit-maximising price. These arrays work as extra dimensions to the landscape, as the information stored in the cell of one array interacts with the other arrays and with information in the agents' side.

For agents, a key set of attributes is their location as (x, y) coordinates, that is used to calculate all distances, essential for any other calculation. They are also used to read and write information from and to the environment arrays. They are also used to plot the charts seen in the following subsections. The locations were plotted with households as the centre of the grid cells, and firms as little pentagons with their colour ranging from red (low profit) to blue (higher profit). Se choose to include a set of hub-and-spoke lines from households to their patronised shops, coloured at random. Other attributes stored for each agent are: income, utility level, consumption levels, rent paid, patronised shop, price paid, behaviour adopted (for households); profit level, practised price, quantity sold, rent paid (for firms). Each household or firm instance also includes an unique ID.

A key aspect of the approach adopted for this simulation is a strictly local optimisation process. This means that any information shown is not calculated *ex ante* and applied to the model, but rather a direct result of local interactions. By this, we intended to avoid bringing any vices from analytical and geometrical implications

(Section 5.4), that use a continuous universe, to this universe of discrete space and time. This implied in extensive iterations that increased computational expenses exponentially. We describe them in the following subsection.

5.5.1 Local optimisation processes

This subsection describes spatially local optimisation processes used to solve the model in the simulation. We describe the processes used for calculating the potential utility for newly entering households, the potential profit for any location-price combination, and any readjustments in patronised shops, quantities, and rents the entrance of a new shop might trigger. We highlight, in advance, that any possible ties between optimum locations is decided by adding a small (order 10^{-6}) random tie-breaker component to the calculated process.

Potential utility calculation

At each time step s, we calculate, for every available location, the potential utility (consumption) achieved for importing the uniform good. Then, we calculate the utility for consuming at every existing shop, highlighting that the newly-entering household will pay the agricultural rent R_A. This yields yet another array with achieved utility values, and a new household will choose that location with the highest utility from the available ones, that will define the fixed utility level $\bar{u}$ that will trigger any rent readjustments in subsequent steps.

Potential profit calculation

For every time step s, we calculate for every location in the landscape the potential profit for a firm entering at that location. We depart from the assumption that shops will marginally outbid any existing households, so that that any location without an existing firm will be a potential location. For each location, we iterate through a range of prices calculating the households that would start consuming there with that specific location-price combination. This generates a total quantity per location, per price, that allows for a potential profit calculation by including the importing fee k and the current rent of the location in the firm's cost function. The optimum price is defined as the price that yields the maximum profit for each location (*argmax*). This informs both the newly-entering firm, that will adopt the profit-optimising price at the highest-profit location, as well as price readjustments for any existing firms, simultaneously defining their prices miopic of the readjustments of the other existing firms by assumption.

Weighted Voronoi diagram

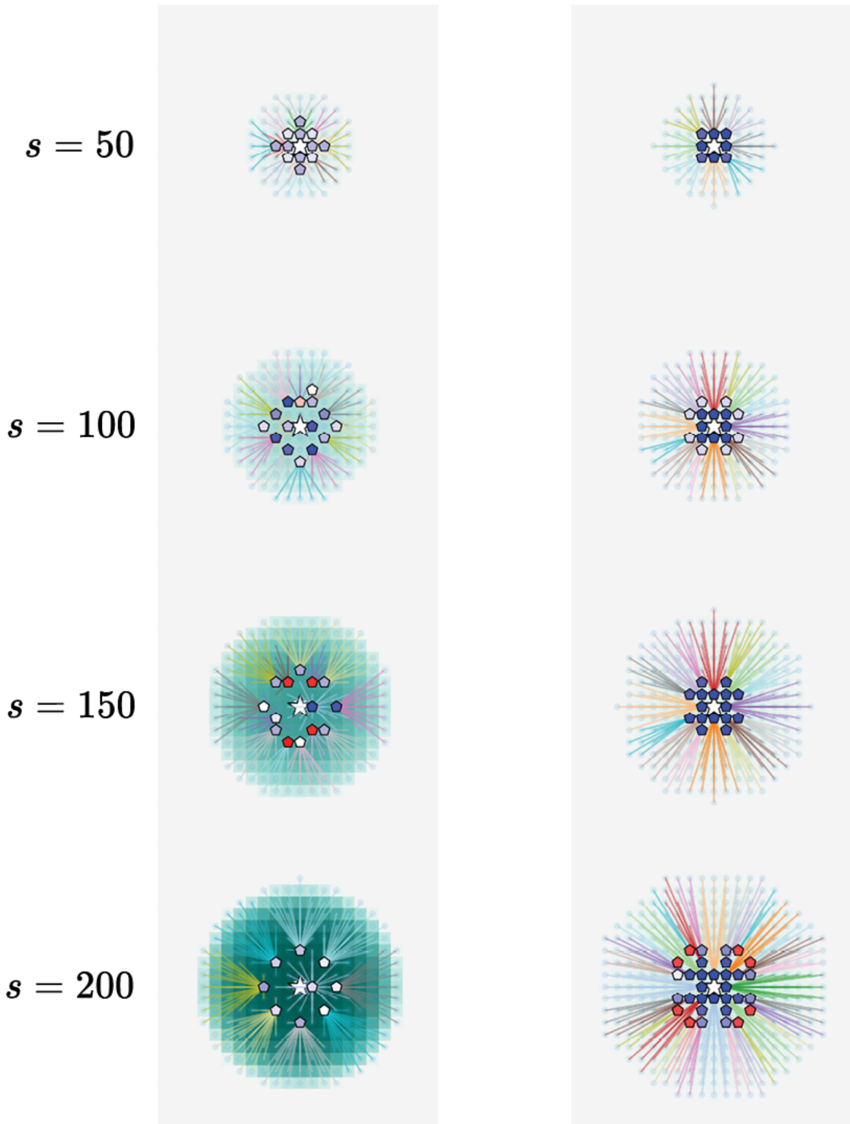
We propose the calculation of what we called a weighted Voronoi diagram, subdividing the landscape among the present firms, effectively calculating their market-catchment area, weighted by their prices and transportation costs. Restricting ourselves to the bid-rent theory, we first calculate the bid-rent of importing the uniform good, that is used as a baseline. Then, we iterate through every firm calculating the residential bid-rent for all locations for consuming at that firm, with that price. Each cell is then assigned to the firm that allowed the highest bid for that location. In parallel, we also store the paid price for that maximisation to happen in a separate array. Since this calculation is performed at the end of every simulation step, it updates any possible changes of consumption quantities and patronised shops a new firm might have originated. Every new information generated here is updated in each agent's attributes.

5.5.2 Benchmark model

We define the benchmark model as the one with a set of parameters that allow us to vary each one of them individually, and highlight which alterations happen to the equilibrium landscape. After a trial-and-error approach, we define the parameters of the benchmark model as: $Y = 30.0$, $k = 5.0$, $R_A = 3.0$, $c = 0.2$, $\gamma = 0.5$. For this range of parameters, we compare the patterns achieved by exponential and linear transportation costs, for time steps $s = 50$, $s = 100$, $s = 150$, $s = 200$. After this, we choose the exponential transportation cost as our baseline model to be tested against the changing of parameters. The results of the simulation for these settings are in Figure 5.5.1.

The results of the benchmark simulations show some patterns already expected by the analytical and geometrical implications. Initially, we have corrected predicted that initial profitable firms will appear near the centre of the city, both for linear or exponential transportation costs. Although slightly different in form, $s = 50$ exhibits a similar equilibrium for both linear and exponential. However, as time

Figure 5.5.1: Benchmark model simulation results for exponential transportation costs (left) and linear transportation costs (right). Equilibrium locations



progresses, we see these patterns differentiating between the two. We highlight that, at $s = 100$, firms have already left the CBD for exponential costs, but not for linear costs. The pattern described of firms "leaving" the CBD exhibits precisely the "moving" dynamic described previously: a newly entering firm will choose a location that minimally differentiates from an existing one, but that triggers the stealing of its clients and the former's demise. This happens over and over, to a point of relative stability observed at $s = 200$, in which 8 firms form a seemingly perfect circle around the CBD, in the form of an octagon, with still one profitable firm very close to the CBD.

For linear transportation costs, we do not visualise any possibility of firms leaving the centre-most location except to limitations of the grid itself. The use of a square-grid in this context of linear transportation costs is a limitation of the simulation. We perceive that firms only prosper when they find a location exactly over a line in-between any household and the CBD. This generates a seemingly *hashtag*-pattern ($\#$), what was not foreseen in analytical implications. Although surprising, it is not completely unanticipated, as we had highlighted in analytical implications that any out-of-the-line location solution for firms in linear transportation costs would be a sub-optimum one. This is, again, a confirmation of expected patterns described previously, although the exact form of this pattern was, indeed, unexpected.

5.5.3 Comparative statics

We depart to explore the comparative statics from the exponential transportation costs benchmark model. We choose it specifically because of the detected unchanged *hashtag*-pattern of the linear transportation costs, whenever firms are viable. There are parameter combinations that firms are, indeed, not viable for the linear configuration, but since our main interest is on the different patterns of firm location, we choose the exponential implementation for varying the other parameters. As highlighted previously, we choose a parameter set for exploration that allows for higher variations in achieved patterns for small variations of parameters. Thus, we choose to test for varying, one parameter at a time, parameter $c = [0, 1]$, parameter $k = [0, 10]$, parameter $Y = [10, 100]$, parameter $R_A = [1, 10]$ and, finally, parameter $\gamma = [0, 1]$. These are depicted in Figure 5.5.2.

We first highlight which parameters act towards changing achieved city size

Figure 5.5.2: Parameter space exploration for exponential transportation cost benchmark model.

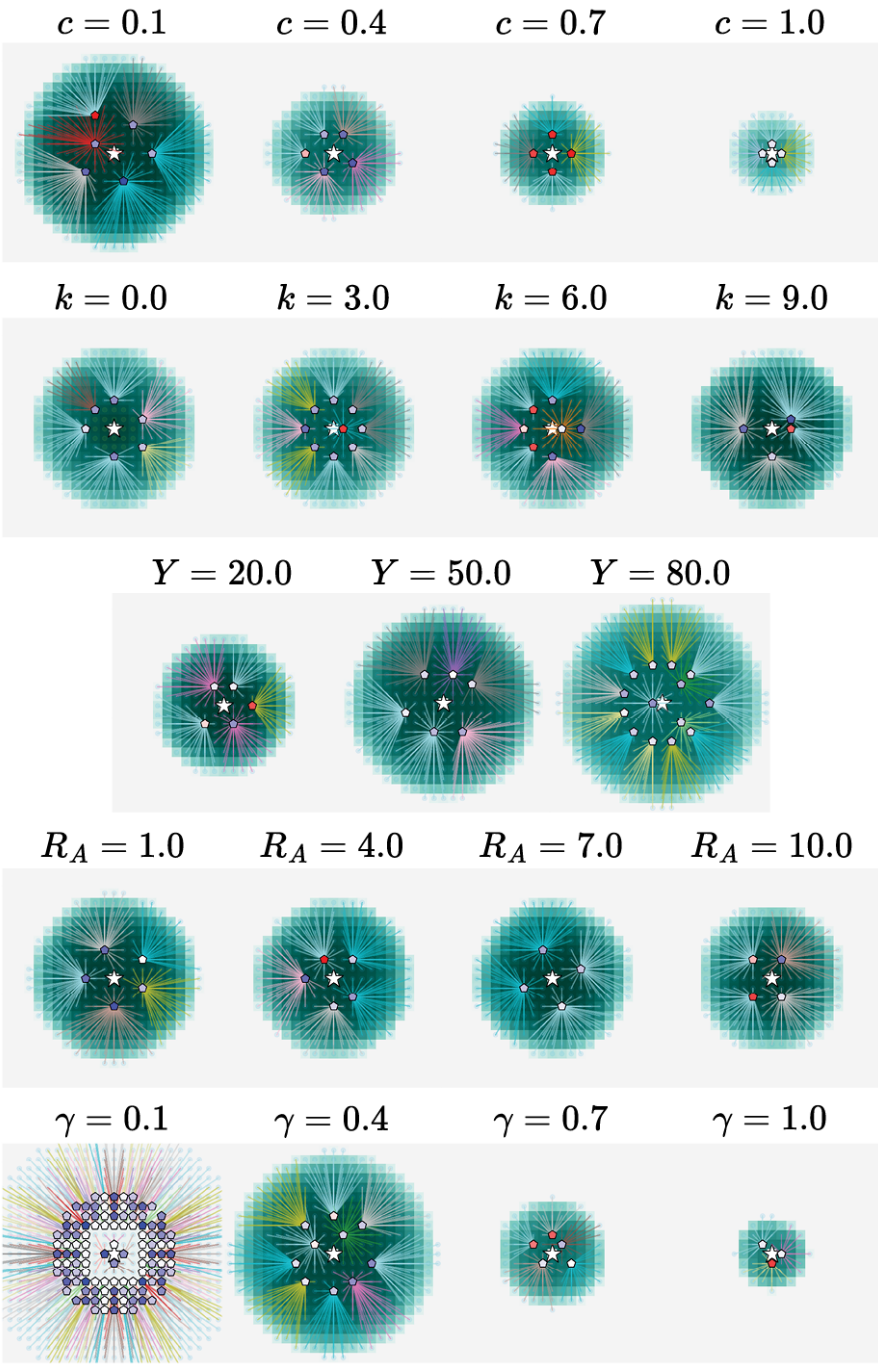
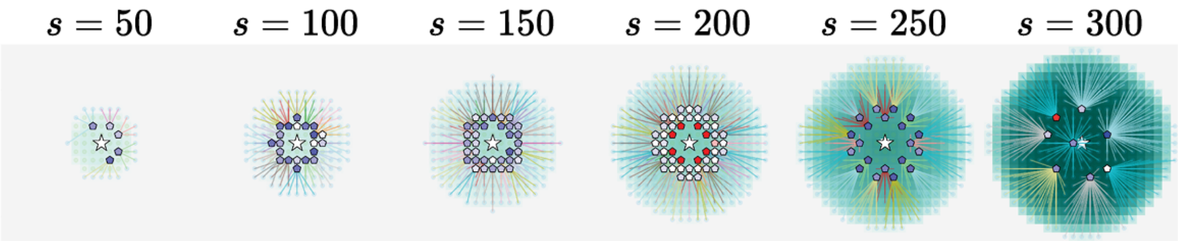


Figure 5.5.3: Time-wise analysis of a high-income variation ($Y = 80$), in which initial central households only import the uniform good.



at equilibrium, and, by extension, city fringe $\bar{x}$ itself. We detect both parameters related to transportation costs as inversely proportional to city size. This is expected, as increased transportation costs at the city fringe diminish the capacity of households to achieve the same utility as that of the outside world (u_0). This is the case for varying both c and γ. We also notice that varying the γ parameter, or the marginal transportation costs, greatly affects the total viability of firms, perhaps as a direct consequence of a smaller city size overall, but as well as higher proportional rents diminishing total consumptions. Another parameter, now with a direct proportionality to city size, is household income Y.

The variation of parameters k and R_A seemingly affect mostly the viability of firms, overall, at equilibrium. More specifically, a higher R_A seems to lead to a lower number of firms viable at equilibrium, as well as a slightly smaller city size, perhaps one affecting the other as well. The variation of k has more intricate implications, as we notice for $k = 0$ all centrally located households are importing the good rather than consuming locally. This is also noticed, although not in equilibrium, in the initial time steps of higher Y. In fact, we notice a very interesting pattern for $Y = 80$, in which initial configurations depict all households importing the good, initial firms serving only peripheral households, until finally central locations become viable for a firm to appear. We highlight these in a separate analysis, depicted in Figure 5.5.3.

These implications, of parameters Y and k directly affecting the land-use patterns, is a confirmation of the theoretical analysis described in Chapter 4. Although with slightly different implementations, we described in Chapter 4 how variations in the relation between income and disposable income net of importing fees greatly affect land-use equilibrium. More specifically, a higher $Y/(Y-k)$ ratio tends to lead central households towards the importing behaviour, inverting the expected curve in which central household consume locally. Since this ratio moves in this direction both for lower k and higher Y, we conclude that the highlighted patterns for Figure 5.5.3 are a direct depiction of this phenomenon.

5.6 Conclusion

This research develops a micro-economic urban model, departing from Alonso's (1964) monocentric-city model and shifting its focus towards differentiating transportation cost functions and the equilibrium location of retail firms. We proposed to expand traditional models in the field, by moving from one dimension (the city-line or circle) to two (discrete locations), and from 2 firms to n, effectively bridging monocentric-city models and location-competition models. We proposed a simplified, consumption-based model, in which households trade-off transportation costs (to a CBD and an adopted shop) and rent (akin to Alonso), innovatively including an outside option of online-shopping. We test for different transportation cost functions (linear and exponential), revisiting a seminal discussion between the minimum and maximum differentiation principles. Finally, we make use of agent-based simulations to account for the complexities, co-dependencies, and non-linearities the combinations between different transportation costs and n-firm competition generate in a two-dimensional landscape, and to solve for both optimum price and location of firms.

We revisit the discussion between minimum vs. maximum differentiation principles for firm location equilibria, whose first exponents were respectively Hotelling (1929) and d'Aspremont et al. (1979). The two concepts are defined, respectively, by linear and quadratic transportation costs, for which we test the prevalence in this research. We anticipated, theoretically, that even the minimum-differentiation principle (in a linear setting) could generate an unexpected pattern of differentiation (called here the "moving dynamic"), once space and time are discretised. This is the case once, by assumption or definition, two firms cannot occupy the same discretised cell and the minimum spatial differentiation actually means two adjacent cells. This also holds for minimum differentiation of optimum prices, once the same logic of discretisation can be applied to the number of decimals of numerical approximations. In the specific case described in this paper, the optimum price for a second firm is the one low enough for it to dominate the consumption decisions of all customers belonging to the incumbent firm, which ends up being the exact same price. Expectedly, however, moving from linear to exponential transportation costs yields optimum location equilibria for firms in out-of-corner solutions, with a range of locations from which households experience the same transportation costs. We conclude then, that both d'Aspremont-like and Hotelling-like local equilibria are possible, which is later confirmed by simulation.

To account for the complexity of involving n-firms, multiple households, and a two-dimensional landscape, we develop an agent-based simulation, in which households and firms interact with each other and with the environment, reading and writing information to and from each other. We correctly confirm expected patterns from the analytical and geometrical analyses by using the simulation, as well as detect unexpected ones. We observed that, even with exponential transportation costs, a minimum-differentiation principle can prevail, with new shops choosing adjacent locations to existing ones. This was described as a "moving dynamic": subsequent shops opening and closing adjacent to each other generating a pattern visually similar to a shop moving through the landscape. When shops are far enough away from each other, though, we observe the opposite phenomenon: a new shop might choose the farthest point from one another, triggering a pattern akin to the maximum-differentiation principle. The discretisation of the space, or more specifically the format of the grid used, generated an unexpected pattern for the linear transportation configuration: the *hashtag*-pattern, in which the impossibility of an equilibrium location out of a direct line between household and CBD creates a very specific visual pattern in the form of a *hasthag* ($\#$).

In terms of methodology, we highlight the fundamental potential of agent-based simulations to solve a location-competition problem in which the number of interacting agents exponentially increases the complexity of the calculations. This is especially the case when dealing with a two-dimensional landscape. And even more so with non-linearities of, for instance, exponential transportation costs. We contribute to the methods in the field by developing a clean approach in *python* of vectorised computations using native *numpy* arrays as our main base. Although agent-based packages do exist in *python* community, the expected complexity of the calculations forced us to approximate the implementation closer to the machine, by vectorising as many processes as possible. Despite their potentials, agent-based simulations are limited by the discretisation of both space and time, as well as decimals considered for numeric approximations, for instance. This is highlighted as a limitation for this sort of approach, and specific configurations of the grid, for instance, should be considered carefully in any implementation.

References

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press.

Anderson, S. P., & Neven, D. J. (1991). Cournot Competition Yields Spatial Agglomeration. *International Economic Review*, *32*(4), 793–808.

Beckers, J., Birkin, M., Clarke, G., Hood, N., Newing, A., & Urquhart, R. (2022). Incorporating E-commerce into Retail Location Models. *Geographical Analysis*, *54*(2), 274–293.

Beckers, J., Weekx, S., Beutels, P., & Verhetsel, A. (2021). COVID-19 and retail: The catalyst for e-commerce in Belgium? *Journal of Retailing and Consumer Services*, *62*, 102645.

Chen, X., & Song, Q. (2021). Two-Player Location-Price Game in a Spoke Market with Linear Transportation Cost. *Discrete Dynamics in Nature and Society*, *2021*, e9971103.

d'Aspremont, C., Gabszewicz, J. J., & Thisse, J.-F. (1979). On Hotelling's "Stability in Competition". *Econometrica*, *47*(5), 1145.

de Vries, J. J., Nijkamp, P., & Rietveld, P. (2004, September 1). Exponential or Power Distance-Decay for Commuting? An Alternative Specification.

Dragone, D., & Lambertini, L. (2020). Equilibrium existence in the Hotelling model with convex production costs. *Regional Science and Urban Economics*, *84*, 103568.

Fain, J. (2023). Should retail stores locate close to a rival? *Journal of Economic Interaction and Coordination*, *18*(1), 129–162.

Feldman, M., Fiat, A., & Obraztsova, S. (2016). Variations on the Hotelling-Downs Model. *Proceedings of the AAAI Conference on Artificial Intelligence*, *30*(1).

Fleckinger, P., & Lafay, T. (2010). Product flexibility and price competition in Hotelling's duopoly. *Mathematical Social Sciences*, *60*(1), 61–68.

Gao, B., Zheng, Y., Huang, J., Department of Economics and Management, Yuncheng university, Yuncheng 044000, China, & Department of Finance, University of International Business and Economics, Beijing 100029, China. (2021). General equilibrium of Bertrand game: A spatial computational approach. *AIMS Mathematics*, *6*(9), 10025–10036.

Guo, W.-C., & Lai, F.-C. (2013). Nesting Horizontal and Vertical Differentiation with Location Choices. *Pacific Economic Review*, *18*(4), 546–556.

Guo, W.-C., & Lai, F.-C. (2015). Spatial Cournot competition in a linear–circular market. *The Annals of Regional Science*, *54*(3), 819–834.

Guo, W.-C., & Lai, F.-C. (2020). Spatial Cournot competition in two intersecting circular markets. *The Annals of Regional Science*, *64*(1), 37–56.

Gupta, B., Lai, F.-C., Pal, D., Sarkar, J., & Yu, C.-M. (2004). Where to locate in a circular city? *International Journal of Industrial Organization*, *22*(6), 759–782.

Hamilton, J. H., Thisse, J.-F., & Weskamp, A. (1989). Spatial discrimination: Bertrand vs. Cournot in a model of location choice. *Regional Science and Urban Economics*, *19*(1), 87–102.

Heywood, J. S., Li, D., & Ye, G. (2022). Mixed duopoly under hotelling with convex production costs. *Annals of Regional Science*, *69*(2), 487–510.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, *39*(153), 41–57.

Irmen, A., & Thisse, J.-F. (1998). Competition in Multi-characteristics Spaces: Hotelling Was Almost Right. *Journal of Economic Theory*, *78*(1), 76–102.

Kohsaka, H. (1986). The Location Process of Central Place System within a Circular City. *Economic Geography*, *62*(3), 254–266.

Larralde, H., Jensen, P., & Edwards, M. (2006). Two dimensional Hotelling model: Analytical results and numerical simulations. *HAL*, 8.

Netz, J. S., & Taylor, B. A. (2002). Maximum or Minimum Differentiation? Location Patterns of Retail Outlets. *Review of Economics and Statistics*, *84*(1), 162–175.

Takahashi, T. (2014). Location competition in an Alonso–Mills–Muth city. *Regional Science and Urban Economics*, *48*, 82–93.

Ushchev, P., Sloev, I., & Thisse, J.-F. (2015). Do we go shopping downtown or in the 'burbs? *Journal of Urban Economics*, *85*, 1–15.

Conclusion

This dissertation offers a comprehensive examination of the spatial distribution of retail establishments and their evolving patterns as cities grow and develop. The findings contribute both empirically and theoretically to the understanding of retail dynamics in urban environments, with a specific focus on how shops cluster, how retail patterns are influenced by city size, and how evolving consumer behaviours, such as online shopping and teleworking, impact the traditional agglomeration forces that have long shaped urban retail landscapes.

In the first chapter, we explore the retail distribution across 782 European cities, highlighting that the quantity, richness, and diversity of retail outlets scale positively with population size. While the total number of shops follows a superlinear elasticity - particularly pronounced in smaller cities - the variety of retail categories scale sublinearly. A nested structure is observed where larger cities encompass most categories found in smaller ones, reaffirming concepts from the Central Place Theory (CPT). The high degree of nestedness (NODF = 0.78) confirms this hierarchical structure of retail composition and introduces a novel application of complex system analysis - in particular, ecological systems - to retail distribution. These findings can have significant implications for urban studies, providing both contributions to understanding how retail ecosystems evolve with city size, as well as how their composition reflects consumer behaviours and centrality in a system of cities.

In the second chapter, we expand the analysis done in the first one, by bringing attention to the intra-urban patterns, more specifically on how accessible these retail outlets are to the inhabitants of these 782 European cities. The study connects retail distribution and local population by the street network, highlighting which areas of the cities are more or less accessible to retail. Findings indicate that retail concentrations follow closely population densities, in the sense that the highest shop densities only occur in densely populated areas - despite trends towards suburban shopping malls. It then aggregates these at the city level, finding that medium-sized, Southern European cities exhibit exceptionally high accessibility within shorter travel times, most likely due to the dense configurations of their urban fabric. Larger urban cores, like Paris, London and Berlin lead the aggregated accessibility over longer thresholds, reflecting the superlinear patterns detected previously. The concept of "tightness" is also introduced, by comparing short- and long-haul accessibilities, and we propose an index of 15-minute tightness to assess this.

One of the key contributions of the dissertation can be drawn by combining the conclusions of the two first chapters: larger cities are home to a disproportionately higher number of shops, but middle-sized cities, particularly in Southern Europe, offer the highest levels of accessibility to retail establishments within the 15-minute city framework. In other words, despite larger cities offering a disproportionately higher number of shops, these are not in general more accessible by their population wihtin a short walking distance. This finding is particularly relevant in the context of recent urban policies that promote proximity and local shopping as a means of enhancing sustainability and livability. The studies show that while larger cities might offer more retail options, it is the middle-sized cities that perform better in terms of accessibility, providing a valuable lesson for urban planners and policymakers aiming to create more inclusive and accessible urban spaces.

In a system of cities, such as the European one analysed in Chapters 1 and 2, we also found that more ubiquitous retail is found in both smaller and larger cities, while more specialised (less common) ones are mostly contained to the latter. This was later confirmed by applying a nestedness metric commonly used in ecological systems. Further work is encouraged to test this for other systems of cities, as to our knowledge no other work assesses this nested structure in a systematic way.

By combining the conclusions discussed above, one important implication arises. Since larger cities both offer more shops and contain shop categories that are not present in smaller ones, while knowing that these are not more accessible by their inhabitants, it is an indication that there is a mismatch between the location of specialised retail and population. Following the hypotheses of earlier works, such as the Central Place Theory (Christaller, 1966), these specialised shops are more likely to be present in more central areas. Due to the larger areas over which larger cities spread their population, it would be a reason why these are not accessible within a shorter travel-time threshold, while appearing in longer ones. Testing whether this holds empirically, while also modelling the explicit mechanisms for this to happen, is one important unfolding to be developed in future work.

When discussing the actual desirability of a 15-minute city pattern, urban planners and theorists should bear in mind the implications of a self-contained community conceptual model. First of all, it is not realistic to think that, in an increasingly connected world, with diminishing transportation and delivery costs, self-contained communities will behave like islands in the urban fabric. Cities may well be a juxtaposition of several 15-minute cities, but, almost by definition, this creates overlays of connectivity that will, necessarily, induce unexpected macropatterns as outcomes. Agglomeration effects will be present and these overlayed subsystems are the very forces that allow for specialised retail to emerge. These are, by no coincidence, traditional centres, still playing a strong role in offering specialised (i.e., less ubiquitous) retail categories to its citizens. Therefore, efforts towards promoting 15-minute cities should still consider the importance of central areas in offering services and shops not commonly observed in more widespread retail centres. This is also valid for when comparing different cities, as larger cities can offer this specialisation more intensely. Perhaps, what we are really observing here is a fractal system of retail centres, ranging from intra-urban to inter-city networks. Considering it as a unified fractal network may allow future work to focus on the importance of different-level nodes, and whether the expected categories are present given the degree of centrality of each node.

The following chapter, derived from previous works but published under this dissertation, tests the transferability of scope and scale of a concept commonly applied to trade at the country level: the Economic Complexity Index. Departing from the assumption of space as an essential mediator of economic activity, we first detect the specialisation of spatial units in specific economic categories. This is an innovative approach as the concept of specialisation is most often applied to countries exporting higher-than-expected volumes of a certain product. We apply it to their mere concentration (derived from accessibility indices) in space. A spatial unit is detected to be specialised in, for instance, furniture, if it has a higher-than-expected concentration of furniture-dedicated shops. Despite deviating from the initial scope of this dissertation, the inclusion of retail in the Urban Economic Space positions this broad category of economic activity within the overall economy of a city.

Another natural unfolding of the work developed in this dissertation, in complementarity with the chapter mentioned in the previous paragraph, is the application of the Economic Complexity methods specifically to retail. Other studies have shown how the index, when applied to retail, can detect the degree of centrality of both areas and categories (Kim et al., 2024). In other words, it may be a way to quantify the level of centrality put forward by (Christaller, 1966). Preliminary results from one of the works unfolding from this dissertation also point in this direction, highlighting retail categories expected to be more prevalent in space, such as bakeries or supermarkets, as "less complex".

The nexus between the Economic Complexity literature and firm networks distributed in space may be key to understanding cities as evolutionary systems, akin to Pumain (2011). Retail is not only a system in itself, it is also connected to other business categories as its specific categories may be influenced or influence the presence of related street-level services, or even office firms. For instance, shops offering fast lunch options may strongly coincide with the presence of law firms or financial institutions, while pharmacies may coincide with healthcare facilities and doctor clinics. These are complementarities expected to arise from the inclusion of multiple economic categories in an analysis, similar to the Urban Economic Space of Chapter 3. In extension to the networks of retail centres described previously, or rather overlaying it, are different networks of other economic activities, which may behave in a similar fashion or not, depending on the nature of the activities. Further work is also encouraged to explore this interconnectivity between different networks of firms.

In the second part of the dissertation, we dive into understanding the dynamics behind the patterns of retail clustering observed in the accessibility analyses, aiming at observing the impact of recent trends (i.e., online-shopping and teleworking) on the form these concentrations take place. More specifically, in the fourth chapter, we develop a theoretical monocentric city model that incorporates online-shopping and teleworking, finding that diminishing delivery costs reduce the viability of central retail and induce urban sprawl, increasing negative externalities to urban dwellers. By making suburban life more attractive, online shopping may undermine central retail by contributing to rising rents and shifting consumer habits. By decreasing visits to the city centres, teleworking reinforces these expected patterns, altering land-use equilibrium especially among higher-skilled individuals (more likely to telework and shop online). The model serves as a theoretical foundation for understanding the impacts of online shopping and teleworking on urban landscapes, hinting at further research that will expand it into two-dimensional, incorporating also suburban retail.

Finally, in the fifth and last chapter of this dissertation, we extend the previously described model by bringing the microeconomic urban framework to two-dimensions, incorporating a $n-$firm game, bridging monocentric-city and location-competition models and innovatively incorporating online shopping. In a consumption-based utility framework, where households trade-off transportation costs (for commuting and shopping trips) and rent - including an external online-shopping option - the study explores different transportation cost functions (linear and exponential) and revisit classical theoretical discussions such as the minimum vs. maximum differentiation principles (d'Aspremont et al., 1979; Hotelling, 1929). Using agent-based simulations, it suggests that both differentiation equilibria may emerge and even coexist in a price-location game, revealing unexepcted dynamics like the "moving", and surprising patterns such as the "hashtag pattern" due to space and time discretisation. The findings highlight the potential of agent-based simulation to solve complex location-competition problems in two dimensions, while acknowledging limitations related to discretisation and numerical approximations.

Despite the limitations of purely theoretical works, they can offer simplified understandings of complex, multi-layered phenomena. The challenging of traditional push-and-pull factors related to consumption and work leads us, as a society, to unknown paths of firm and population location patterns. The theoretical works of this dissertation highlight how these can impact traditional centres, and especially brick-and-mortar retail in these areas. Given the historical importance of retail spaces, serving not only as places of consumption, but as places of exchange of experiences and connectivity between city dwellers, it is vital to try to preserve them. The continued centrality of retail hubs for specialized products suggests that central locations still play a crucial role in urban consumption, even as consumer behaviours shift. The results indicate that while these emerging trends do threaten the traditional role of city centres, the resilience of local governments and the strategic responses of retail businesses will play a determining role in how these forces reshape urban retail.

The application of an agent-based model is a significant methodological contribution of this dissertation. By innovatively including online-shopping in a model of retail location, it bridges quantitative geography and urban economics literature, offering a tool for solving complex location equilibria problems with multiple firms in a two-dimensional landscape that urban economics falls short of addressing in analytical terms. It also contributes to the theoretical discussion between minimum and maximum differentiation principles, historically dealt in location-competition models, by proposing that these two may coexist in a dynamic, complex system. The model, overall, provides a more dynamic understanding of how retail firms choose their locations and how these choices are influenced by both physical and digital marketplaces. Both the theoretical and the simulation models offer a framework to include the relatively new force of online-shopping in traditional urban modelling approaches.

In conclusion, this dissertation provides a thorough analysis of urban retail location, both in a theoretical and an empirical sense, offering valuable insights into both the realities observed and theoretical implications of retail location dynamics. As cities continue to grow and evolve, the findings presented here are presented to the academic community and to urban planners at large as guidance for future urban policies, particularly those focused on promoting sustainable development, local accessibility, and resilient retail systems in the face of changing consumption patterns.

References

Christaller, W. (1966). *Central Places in Southern Germany*. Prentice-Hall.

d'Aspremont, C., Gabszewicz, J. J., & Thisse, J.-F. (1979). On Hotelling's "Stability in Competition". *Econometrica*, *47*(5), 1145.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, *39*(153), 41–57.

Kim, J., Yu, D., Choi, H., Seo, D., & Jun, B. (2024, July 29). Redefining Urban Centrality: Integrating Economic Complexity Indices into Central Place Theory.

Pumain, D. (2011). Systems of Cities and Levels of Organisation. In P. Bourgine & A. Lesne (Eds.), *Morphogenesis: Origins of Patterns and Shapes* (pp. 225–249). Springer.